

Passing the torch

With the death of Francis Crick, biology is mourning one of its deepest thinkers. A work of futurology, published in 1970, reveals the extent of his prescience — and suggests challenges for today's theorists.

When Francis Crick and James Watson described the double-helix structure of DNA in 1953, they charmed many with their unassuming pay-off line: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

In fact, little in the broad landscape of biology escaped Crick's attention. The publication of his obituary (see page 845) marks a good time to revisit one of his lesser-known works, a 1970 essay in *Nature* on 'Molecular Biology in the Year 2000' (www.nature.com/nature/focus/crick/pdf/crick228.pdf). Unlike most exercises in crystal-ball gazing, it holds up remarkably well to the scrutiny of hindsight.

Setting about his task with some trepidation, Crick divided biological problems into three categories: those that would be solved by the turn of the millennium; those that would have advanced considerably towards resolution; and the big questions that would still be taxing the best scientific minds.

In some respects, Crick overestimated the rate of progress. We are still some way from a complete understanding of the significance of repetitive 'junk' DNA; much the same can be said for the sequences that regulate gene expression. But, as he predicted, the replication of DNA and the structure of chromosomes are now well documented.

Crick's choices for the second category were similarly prescient. Some problems fell more easily than he expected: apart from minor details, the mechanism of muscle contraction can be considered solved. Others, such as "the nature of the influences which produce 'gradients' in embryological development", have proved more complex. But his list represents a remarkably accurate summary of the issues that have occupied biologists over the past three-and-a-half decades — not all of which would have been obvious back in 1970.

As for the final category of unsolved problems, we're still wrestling with the origins of life on Earth, and whether life exists on other

worlds. The nature of consciousness — to which Crick devoted his later years — similarly remains shrouded in mystery.

Crick also had some observations about the future of this journal. "I hope," he wrote, "that *Nature* will not become overloaded with too many detailed papers in what we now regard as the more classical parts of molecular biology."

In terms of what was then seen as "classical", Crick's wish has been granted. But in recent years, there's no denying the preponderance of *Nature* papers on the molecular details of such topics as signal transduction and the cell cycle. Crick would probably have been disappointed by the relative paucity of great contributions on the origins of life, and astrobiology. In part, this may reflect a conservative system of research funding that discourages biologists from attempting the high-risk imaginative leaps that Crick favoured. But it's also a reflection of the nature of the beast. Crick's essay hinted at this: "Problems involving complex interactions can hardly be avoided, since some of the most profound aspects of biology are of this character."

Crick understandably failed to predict the rise of high-throughput techniques, including whole-genome sequencing, DNA microarray analysis and proteome profiling. These techniques have confirmed, if there was any doubt, that the details of biology are truly devilish.

In the wake of genomics, the emerging discipline of systems biology has the potential to make sense of these details, by treating genes and proteins as networks, ripe for theoretical analysis. There's reason to hope that theorists will soon be able to advise lab-based biologists on where to look for answers, much as theoretical physicists predicted the existence of fundamental subatomic particles a generation ago.

Molecular biology is easily dismissed as a book of recipes, and too many of its practitioners allow this criticism to pass unchallenged. Crick would not have wanted this to happen. Let's hope his legacy of thinking big lives on. ■

Let's blame Canada

Americans should worry less about their neighbour and more about the prestige of regulators who protect public health.

Lester Crawford, acting commissioner of the US Food and Drug Administration (FDA), is worried that the importation of cheap pharmaceuticals could expose Americans to attack from bioterrorists. In an interview with the Associated Press on 11 August, he raised the potential contamination of imported drugs as a threat to national security.

In practice, the imports come mainly from Canada, where drugs are subject to government price controls. They are increasingly sanctioned by state governments in the United States, where citizens are tiring of paying inflated prices for medicines. The Bush administration — taking its cue from US drug companies — would like to close the door to such imports.

As that acclaimed documentary the *South Park* movie demonstrated, the United States' paranoia about its threatening northern neighbour is richly justified. But Canadian perfidy must not blind Americans to a starker domestic threat: the degeneration of once-

prestigious federal agencies, such as the FDA, into political poodles.

The FDA has an inspiring history. Its record of stalwart independence is one of the reasons Americans have such high confidence in the safety of both their pharmaceuticals and their food supply. Only a few years ago, for example, FDA commissioner David Kessler launched an audacious effort to regulate tobacco as a drug. Congress opposed the move and Kessler received little support from his bosses in the Clinton administration. But his stand, for a while, had 'big tobacco' shaking in its boots.

Today, the chances of such an initiative originating within the FDA itself are slight — not much larger than the chances of someone choosing to terrorize the US public by contaminating a shipment of drugs before their export from Canada. It is scandalous that the FDA's leadership seems ready to make politically motivated pronouncements that link two serious issues — drug pricing and bioterrorism — in a manner likely only to inflate public cynicism about both. ■





Valley life
Climate model offers
California a taste of
things to come
p818



Seeing red
Marine scientists
bid to halt ocean
acidification
p820



Desert storm
UK inquiry seeks
clarity on 'Gulf War
syndrome'
p821



Viral strain
Vietnam confirms
three human deaths
from bird flu
p822

Biologists fear cloning hype will undermine stem-cell research

Jonathan Knight

A dose of reality needs to be injected into the excitement surrounding therapeutic cloning, senior stem-cell biologists have warned.

Researchers fear that optimism generated by recent advances, including the award on 11 August of the first UK licence for research on the technique, has raised expectations of individualized cures for degenerative diseases. In reality, say those in the field, such a prospect remains distant at best.

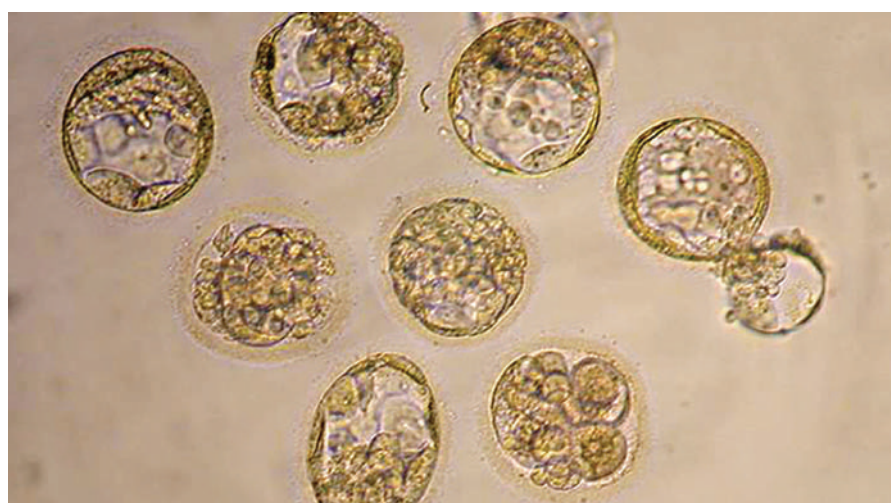
"If we keep talking about cures and they don't come in the next few years, that's going to be a problem," says Jose Cibelli, a stem-cell researcher at Michigan State University in East Lansing.

The decision by the Human Fertilisation and Embryology Authority to grant a licence to the Newcastle Fertility Centre, part of the city's Life biotechnology village, triggered a storm of media attention. The licence allows scientists to create human embryos by inserting nuclei from human skin cells into human eggs. The Newcastle team will then be able to extract stem cells from the cloned embryo.

Although researchers such as Cibelli were thrilled at the news, some worry that the media coverage underplays just how few labs are working on therapeutic cloning. They also say that short-term benefits of the work, such as providing fresh insight into diseases such as diabetes and Parkinson's (see *Nature* **422**, 787; 2003), are not acknowledged.

Short-term and long-term benefits are usually mentioned in the same breath, but the possibility of cures seems to overshadow the research benefits, says Lawrence Goldstein, a stem-cell researcher at the University of California, San Diego. "Many of us in the scientific community talk about both of these, but the press often seizes on one, as do the public and the patient advocates," he says.

Talk of cures is justified, say experts, but only in the long term. Researchers believe that embryonic stem cells can be coaxed to form any tissue in the body. So if they can be extracted from embryos cloned from a patient, they could be used as genetically identical replacements for damaged tissue. This would avoid the risk of rejection associated with transplants.



Media attention on these cloned embryos, created this year, has obscured short-term research benefits.

Yet the scientific hurdles that remain are immense. Cloning is tremendously inefficient, for example. When researchers at Seoul National University in South Korea produced the world's first human embryonic stem-cell line from a cloned embryo in February, they got through 242 eggs from 16 women. Until that success rate is improved, cures or treatments from therapeutic cloning will be impossible. "There is a limited number of eggs for all the patients in the world," says Cibelli.

Limited company

Efficiency may improve with time, although that will require repeat trials with larger numbers of eggs. But sources estimate that fewer than five labs around the world are even trying. The South Korean group has halted its cloning work until a new licensing law comes into effect next January (see *Nature* **429**, 12–14; 2004). Singapore, home to several stem-cell groups, has not yet finalized legislation relating to the licensing of therapeutic cloning.

US law prohibits federal funds from being used for therapeutic cloning, and private labs are not thought to be working on the technique. And in Israel, another country with stem-cell expertise, researchers are discouraged because only a small number of eggs can be used in experiments, says a US scientist

who has collaborated with colleagues there.

The Newcastle group faces fewer problems, as researchers there will have access to eggs left over from *in vitro* fertilization procedures. These number some 2,000 a year, says Alison Murdoch, a reproductive biologist with the group, although the actual number available will depend on how many women give permission for their eggs to be used.

Murdoch and her colleagues will initially attempt to generate a cell line from the cloned embryo of a healthy subject, to make sure the procedure works. Only then will they try to produce a cell line from a patient with diabetes, the clinic's primary disease focus.

Even getting the first healthy cell line will be hard. The South Korean team used adult cell nuclei transferred from cumulus cells surrounding the donor's egg. Using nuclei from the donor avoids problems with mitochondria — energy-producing parts of a cell that carry their own genetic content, which is not always compatible with any mitochondrial DNA found in the donor's egg. Cumulus cells also combine relatively easily with eggs.

The Newcastle group, by contrast, will take nuclei from skin cells and attempt to fuse them with eggs from many different donors. "If we can get the technique to work reliably in the first year we will have made a great deal of progress," says Murdoch. ■

Crisis fomented as unstable lake builds in the Himalayas

K. S. Jayaraman, New Delhi

Landslides on the India–China border have created a lake holding millions of cubic metres of water. Scientists and engineers in the region are puzzling over how to avoid a catastrophic collapse of the lake wall, but Indian researchers say their plans have been hampered by China's refusal to allow them to visit the site.

The Chinese government warned on 11 August that landslides had blocked the Pare Chu river in Tibet, creating a lake about 40 metres deep and holding 60 million cubic metres of water. China said it was evacuating people downstream and advised India to do the same. The river enters India about 35 kilometres from the lake and, as the Sutlej, continues past several towns in Himachal Pradesh.

According to Muthaia Perumal, a hydrologist at the Indian Institute of Technology (IIT) in Roorkee, the water will rush into a narrow gorge should the lake wall break. "It will not be like floods in the plains where water can spread out," he says. "We expect a huge wall of water moving in bulk without its height diminishing."

Creating tunnels in the lake wall could have provided an exit for the accumulating water had measures been taken when the landslides occurred in early July, says Manoj Datta, a civil engineer at the IIT in Delhi. "Now the dam is overflowing it is too risky to try anything," he adds. The Indian government has evacuated 50,000 people from the danger zone.

Indian researchers regret that a joint strategy was not put in place early on, but the Indian government said on 13 August that a request to China for permission to send scientists and engineers to the site had been rebuffed. The Chinese government says the site is difficult to reach and is briefing Indian authorities on progress made by its own team at the lake. Chinese hydrologists working on the lake could not be reached for comment. ■



Dam danger: a catastrophic flood is feared after landslides blocked a Tibetan river.



Green remembered hills: the wine of the Napa valley could lose its edge if emissions are not reduced.

Climate modellers go local to target California's politicians

Emma Marris

A "pioneering" example of a new breed of climate model, designed to plug the gap between academic research and political decision making, has been published.

The analysis of California's climate by Katharine Hayhoe and colleagues outlines how conditions in the state will change under two alternative scenarios: one models the state's future under aggressive policies to reduce carbon dioxide emissions; the other predicts its fate without them. The result is an analysis of the implications for heat-related human mortality, the water cycle and agriculture. It is designed to influence the state's policy-makers.

"Other researchers looked at just one scenario, which is kind of fatalistic," says Hayhoe, who is based at the University of Illinois at Urbana–Champaign. "This way you have a choice."

Rajendra Pachauri, chairman of the Intergovernmental Panel on Climate Change (IPCC), has flagged up such regional studies as a vital part of the panel's next assessment (see *Nature* 417, 106; 2002). In this case, the project was suggested to Hayhoe by the Washington-based Union of Concerned Scientists (UCS), which fears that the findings of climate-change science are not being disseminated in a useful form.

The study, which is published in print on 16 August (K. Hayhoe *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0404500101), uses data from two global climate models, along with local weather histories. In the high-emissions scenario, there will be six to eight times more heat waves and five to seven times more heat-related deaths in 2100 than there are now. Alpine forests will practically disappear, snow packs will shrink and the vineyards will move from producing gourmet wine to plonk.

"What we did is a model for a regional

assessment," says Hayhoe, who adds that the team chose California because the state is both tough and potentially responsive. "It has a lot of different climate zones, so it is challenging. Also, California is a leader in terms of reduction of emissions. It might actually use our findings."

Hayhoe hopes her forecasts will be politically effective, but says she wants the science to be unbiased and to be perceived as being so. She adds that her team used low estimates for emissions and their effect on the climate: "We tried to err on the conservative side." She also says that after the initial suggestion, the UCS stepped aside, although two authors do list the UCS as an affiliation. State officials declined requests to comment on the paper.

The study is "pioneering," according to Warren Washington, a climate researcher at the US National Center for Atmospheric Research in Boulder, Colorado, who provided tailored data from the centre's global climate model. He predicts that the format will be popular. "This type of paper and analysis will be repeated for many regions around the world," says Washington.

Regional assessments are becoming more feasible as computing power improves, adds David Viner, of the IPCC's Data Distribution Centre at the University of East Anglia in Norwich, UK. The regional models, although huge, can now be done on off-the-shelf computers. "I've got a G5 Power Mac on my desk, which has the capability of a supercomputer from five or ten years ago," says Viner. "There are lots of people doing this now."

He adds that moving the data around is probably more problematic than producing it. Global models, which can require up to 5,000 gigabytes of memory, are starting to become publicly available on the Internet. However, they do not yet provide the detail needed for a rigorous local model. ■

Newton's religious screeds get online airing

Geoff Brumfiel

"Among the Beasts that represent Kingdoms I reckon the Dragon one. A Dragon signifies the person of a hostile King & serpents according to their bigness the persons of other greater or lesser enemies." These words were penned not by an obscure mystic, but by a man many consider to be the father of modern science — Isaac Newton. They are part of the introduction to the 300,000-word interpretation of the book of Revelation that Newton wrote in the late seventeenth century. The work was published for the first time last month.

The writings reveal a religious fervour that until now has been seen only by scholars of science with access to archival manuscripts, according to Robert Iliffe, a science historian at Imperial College London. Iliffe directs the Newton Project, which posted the documents online on 15 July. "This is the first time that people can see what he really believed," he says.

Historians of science have long known of Newton's interest in religion and alchemy, but few realize the radical nature of his work, says Iliffe. "What he believed would have been hideous to virtually everybody in Oxford and Cambridge for decades to come," he says.



Revelation: the majority of Newton's writings were theological tracts.

Among the unorthodox ideas in the text is a belief that Catholics, led by the Pope, are the false idolaters spoken of in the book of Revelation. Antagonism towards Catholics was not uncommon in Newton's period: during his lifetime England had experienced a bloody civil war that pitted Charles I, widely regarded as a Catholic sympathizer, against the mainly

Protestant parliament, before Charles II was restored to the throne.

But the vehemence of Newton's arguments pushed him well beyond the Anglican faith, to which he ostensibly belonged. "He believed that the Pope was the personification of the Antichrist here on Earth," Iliffe says.

Newton knew such beliefs lay outside the social mores of his peers, and there is no evidence that he ever published any of the writings now available online. Still, Iliffe says, Newton's religious writings constitute more than half of his entire written work.

In the past, many thought that Newton pursued religion only in his spare time, or that the majority of his religious work had been copied from others. But Iliffe claims that these writings show his theological work was carefully planned and often related to his work in mathematics and physics. For example, he sets up his text on the Apocalypse with mathematical formalism, outlining rules, definitions and a proof of his beliefs.

Ultimately, Newton's religion and science may have been tied together by belief in absolute truth. Newton used testable hypotheses to find truth in nature, and believed that his religious writings revealed the truth about God, says Iliffe.

www.newtonproject.ic.ac.uk

Firm sets sights on gene silencing to protect vision

Erika Check, Washington

A much-hyped technology known as RNA interference (RNAi) has moved a step closer to the clinic.

Biotechnology firm Acuity Pharmaceuticals of Philadelphia, Pennsylvania, asked for permission on 10 August to use the technique to treat a common cause of blindness. The clinical trial, if approved by the US Food and Drug Administration, would be the first of its kind.

RNAi, which uses short lengths of genetic material to selectively shut off genes, was demonstrated in human cells in 2001 and has yet to be tested in people. Some studies have raised concerns about just how selective the technique is, so observers say the filing is a landmark for the field but remain cautious.

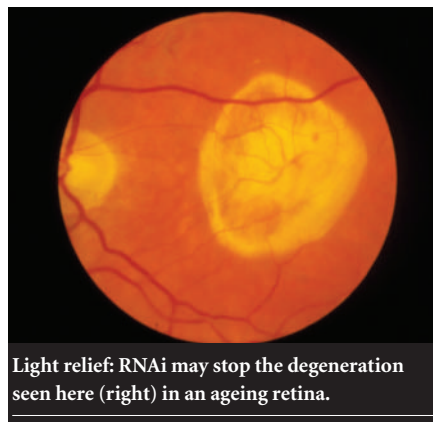
"This is a significant milestone," says Greg Jensen, a biotechnology analyst at Ernst & Young in Palo Alto, California. "Whether it will work out, or give us the first bad news about RNA interference, we don't know yet."

Acuity hopes to use a small interfering RNA to treat patients with wet age-related macular degeneration. The condition is caused by extended growth of blood vessels in the retina — a problem that Acuity thinks can be tackled by silencing the gene that triggers the growth. The company says that the condition affects more than

1.65 million people in the United States.

Dale Pfof, Acuity's chief executive, says his firm has applied for a patent on the RNA molecule involved, but the intellectual-property issues are not straightforward. Rival firm Alnylam, based in Cambridge, Massachusetts, has already requested a patent on the method of RNA delivery that Acuity will use. Alnylam says it plans to begin its own clinical trials in macular degeneration next year.

Pfof is confident that Acuity's disease-specific patent will be granted, but Alnylam begs to differ. "We're very pleased to see silencing RNAs entering the clinical stage," says John Maraganore, chief executive of Alnylam. "But at the end of the day, we believe that anybody developing RNAi therapies needs to talk to us."



Light relief: RNAi may stop the degeneration seen here (right) in an ageing retina.

For more news and analysis go to
news @ nature.com
www.nature.com/news

Kerry pledges to axe Yucca Mountain nuclear-waste dump

Geoff Brumfiel, Washington

Plans for a multimillion-dollar nuclear-waste dump, already shaken by a string of budgetary and legal setbacks, have taken a further knock. The Democrat presidential candidate John Kerry pledged last week to block the scheme if he is elected.

Debate over the Yucca Mountain repository has injected some science into the presidential race, as both candidates campaign heavily for the swing state of Nevada, home to the site. Advocates of nuclear power hope to make Yucca Mountain the nation's primary nuclear dump, but opposition from the state government and scientific questions about the site have delayed the project (see *Nature* 412, 850–852; 2001).

On 10 August, Kerry used scientific studies to claim the plan was flawed, including one saying that corrosion could destroy the metal waste-storage canisters. "I can sum up my stance on Yucca Mountain in four words," Kerry told supporters in Nevada. "Not on my watch."

His attack was rebuffed by President George W. Bush, who two days later accused Kerry of "trying to turn Yucca Mountain into a political poker chip". Kerry may also be out of step with the scientific evidence. According to a 28 July letter from the Nuclear Waste Technical Review Board, the government agency that raised the corrosion concern, new data show that corrosion is "unlikely".

The campaign rhetoric comes at a turbulent time for the site. The facility is designed to last 10,000 years, but on 10 July a federal appeal court ruled that it must comply with a 1995 National Academy of Sciences study that raises this to 100,000 years. The decision came as federal officials were rushing to file a licence application for the site with the Nuclear Regulatory Commission, which oversees nuclear-waste disposal.

Yucca's advocates are also faced with a budgeting miscalculation that led the House of Representatives in June to approve just \$130 million of the \$880 million requested for the site in the next financial year. Many doubt that the bill, when it is completed, will provide full funding. "The project is caught in a perfect storm right now," says Bob Loux, director of the State of Nevada Agency for Nuclear Projects, which opposes the site.

The fate of the site may now depend on the outcome of the presidential election. If Bush wins, he may be able to negotiate a new 2005 budget for the project. ■



Sea change: falling pH levels caused by carbon dioxide emissions could devastate oceanic ecosystems.

Researchers seek to turn the tide on problem of acid seas

Quirin Schiermeier, Munich

Researchers met last week to map out plans to study a serious but largely neglected environmental problem — the gradual acidification of the oceans.

Since the industrial revolution, sea surface pH levels have dropped by around 0.1 units as the oceans absorb atmospheric carbon dioxide. This is already enough to trouble some marine species, but researchers warn that values could fall by a further 0.5 units by 2100. If steps are not taken to cut CO₂ emissions, pH could drop even further, perhaps to levels that are thought to have triggered catastrophic extinction events in Earth's history (*Nature* 425, 365; 2003).

"We're taking a huge risk," says Ulf Riebesell, a marine biologist at the Leibniz Institute of Marine Sciences in Kiel, Germany. "Chemical ocean conditions 100 years from now will probably have no equivalent in the geological past, and key organisms may have no mechanisms to adapt to the change."

Aspects of the threat are already being studied, but marine scientists met at the Plymouth Marine Laboratory, UK, on 11–13 August to start work on a comprehensive research plan. The scheme, to be finalized later this year, will list key scientific questions and provide a blueprint for funding agencies and researchers.

Their efforts were due to be boosted on 17 August by news that Britain's Royal Society is to probe the likely changes to marine ecosystems following a rise in ocean acidity.

Studies of the effects of acidification on

marine organisms will play a prominent part in the research plan. Last month, researchers showed that the shells and hard skeletons of plankton and corals will begin to dissolve as the oceans become more acid (R. A. Feely *et al. Science* 305, 362–366; 2004). The loss of these creatures would have incalculable consequences for the entire marine food chain.

Carol Turley, a senior scientist at the Plymouth lab, says researchers will investigate the growth, reproduction and adaptation capability of marine species ranging from bacteria to vertebrates. Results from these experiments will help to shape protocols for open-ocean studies. With the help of measurement tools such as pH sensors attached to profiling floats, researchers will try to track the subtle ecosystem interactions between the sea surface and the sea floor, and between coastal areas and the open ocean.

Modelling the likely changes in ocean chemistry, and determining how they may affect temperature, salinity and nutrient availability, will be another priority, says Turley. Researchers will also add CO₂ to the ocean to simulate long-term changes. Together, these studies should show how biogeochemistry, species diversity and evolution will change in an acidic ocean.

Riebesell hopes that the initiatives, which will be overseen by the Integrated Marine Biogeochemistry and Ecosystem Research project, will influence funding agencies. "We know quite a lot about the ocean of the past," he says. "But we owe it to people to tell them more about the ocean of the future." ■

Sick veterans pin hopes on Gulf War inquiry

David Osumi-Sutherland, London

How should an illness with no clearly defined symptoms or cause be dealt with? That's the challenge emerging from a UK inquiry into the illnesses facing veterans of the first Gulf War. Soldiers who fought in the conflict are falling sick in surprisingly high numbers, researchers have told the inquiry. Yet experts say there is no definable 'Gulf War syndrome' nor any obvious reason for the symptoms.

The inquiry has been running in London since 6 July, in response to requests by former service personnel. Organized by Alf Morris, a Labour member of the House of Lords who is sympathetic to the veterans' case, it does not have government backing. But former senior army personnel and former employees of the Ministry of Defence have given evidence.

Several large-scale epidemiological studies have been run on veterans of the 1990–91 war and all have revealed significant increases in disease. A study of almost 10,000 UK veterans, for example, found that 24% suffered health problems (N. Cherry *et al. Occup. Environ. Med.* **58**, 291–298; 2001), 14% higher than in a control group of some 5,000 who did not serve in the Gulf.

A wide range of symptoms, from headaches and diarrhoea to memory loss, were reported, ruling out post-traumatic stress disorder. But this diversity made it impossible for the survey's lead author, Nicola Cherry of the University of Alberta in Edmonton, Canada, to distinguish a specific group of symptoms to characterize the condition.

Identifying a cause for the high rates of ill-health has proved just as difficult. Experts at the inquiry have focused on two possibilities: exposure to nerve gas, and the medication given to soldiers before the war.

Robert Haley of the Southwestern Medi-



Call to arms: could vaccinations be the cause of increased illness among British Gulf War veterans?

cal School at Dallas, Texas, cited US government studies estimating that up to 100,000 US troops could have been exposed to low levels of the nerve gas sarin when an ammunition dump at Khamisiyah in southern Iraq was bombed. A study revealed changes in the brains of sick veterans, which he says are consistent with low-level sarin poisoning (R. W. Haley *et al. Radiology* **215**, 807–817; 2000).

Other witnesses dispute Haley's hypothesis. "Sarin is a deadly substance," says Simon Wessely, a psychiatrist at King's College London. Any exposure, he argues, would have resulted in deaths, but none were recorded.

Wessely adds that epidemiological studies show no link between illness levels and the area of the Gulf where the soldiers served, suggesting that researchers should focus on a factor that applied to all veterans. One possibility, he says, is pyridostigmine bromide tablets, taken regularly by troops to protect against nerve-gas attack. This is hard to

study, as records of its use were not kept. But Wessely says that troops given multiple vaccinations in the Gulf are more likely to be ill than those who had the same shots at home. He speculates that the risk factor may be a combination of stress and the vaccinations.

The Ministry of Defence denies that Gulf War syndrome exists. Some 2,500 war pensions have been awarded, but that figure includes payments for all injuries sustained in the conflict. Yet if Cherry's analysis is extrapolated to the 45,000 troops who fought in the Gulf, more than 6,000 extra UK soldiers could have become ill because of the war.

"Veterans are finding it hard to get the medical treatment that they need," says Terry English, director of welfare at the Royal British Legion, which campaigns on behalf of ex-servicepeople. "This should be given a high priority."

The inquiry, chaired by former law lord Anthony Lloyd, is due to end in October. ■

Plan for light relay sparks heated opposition

Mark Peplow, London

The idea was to "unite all nations by the enlightening power of physics". But with eight months to go, Max Lippitsch's plans have succeeded only in dividing his colleagues.

Lippitsch, a physicist at the Karl-Franzens University in Graz, Austria, wants to shine a relay of lights around the Earth to celebrate the World Year of Physics 2005. But astronomers say the scheme sanctions light pollution, the needless illumination of the night sky that increasingly interferes with their observations.

Britain's Institute of Physics pulled out of the project late last month. "Astronomers

were very concerned," says Caitlin Watson, who is managing the institute's contribution to the year's celebrations. "Even though this won't produce much light pollution, it sets a very bad precedent."

The World Year of Physics celebrates the centenary of Albert Einstein's "miraculous year", which saw him lay the foundations for the theory of relativity, quantum theory and the theory of brownian motion. The goal is to boost the public awareness of physics.

The light relay is set to begin on 18 April next year in Princeton, New Jersey, the 50th anniversary of Einstein's death. Participants will shine torches to neighbouring groups up to 10 kilometres away, who will then

switch on their lights. This domino effect should ripple around the Earth, with undersea fibre-optic cables linking continents. Lippitsch says that the light pollution will be tiny because high-powered searchlights and lasers are banned.

But Vinaya Sathyasheelappa, the World Year of Physics project coordinator for the American Physical Society, says there are "mixed emotions" about the project in the United States. He adds that the society is nonetheless likely to back it: "Without the United States the light would have to go through Canada, and Canada's physics society just doesn't have the resources to make this work." ■

Did extra label cause the scare that shut down Los Alamos?

Washington A pair of 'missing' computer disks that sparked a security scandal at the Los Alamos National Laboratory in New Mexico may never have existed, according to a senator who oversees the laboratory.

The belief that disks of classified data had gone missing triggered the shut-down of the nuclear weapons laboratory last month and led its director, Pete Nanos, to criticize the "cowboy culture" of scientists working there (see *Nature* 430, 387; 2004).

But Senator Pete Domenici (Republican, New Mexico), who chairs the Senate committee that oversees the lab, says he now suspects a clerical error. Sources at the lab have suggested that security-coded labels were made up for disks that were never actually created.

"It may be what we have here is a false positive," Domenici said on 10 August. "But the entire situation only reinforces that we need to improve the inventory system."

Laboratory spokesman Kevin Roark declined to comment, citing an ongoing investigation into the incident.



Bird bath: a health worker disinfects ducks arriving at a poultry market in Hanoi.

More avian flu deaths confirmed in Vietnam

Tokyo Avian influenza was the cause of three human deaths from a respiratory condition early this month, Vietnam has confirmed. But World Health Organization (WHO) officials have struggled to get the samples they need to characterize the viral strain involved and to judge the danger it may pose to public health.

On 16 August, Nguyen Thi Kim Tien, director of the Pasteur Institute in Ho Chi Minh City, confirmed that the sample from one victim contained the H5N1 virus — the same strain that killed 15 people in Vietnam and 8 in Thailand in February.

Peter Horby, a WHO epidemiologist in charge of infectious-disease surveillance in Vietnam, is not surprised to see sporadic human cases re-emerging. "It's what we would expect," he says.

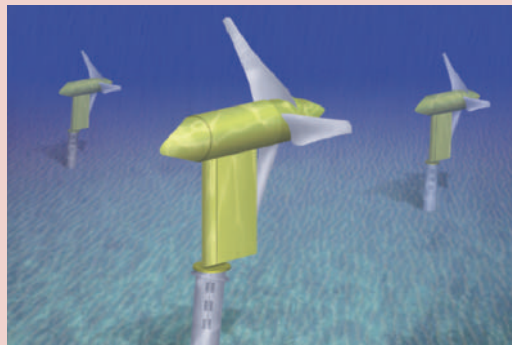
Still, Horby says it is essential that the

Good tidings for alternative power in New York

New York Next month, the fishy inhabitants of New York's East River will acquire some sleek new neighbours. Verdant Power, an energy company based in Arlington, Virginia, intends to plunge six electricity turbines into the river, creating the United States' first tidal-power plant.

The plan is to attach the machines to concrete piles hammered into the bedrock (see right). As the tide surges in and out, the heads will pivot to face the current and the blades will spin. The turbines will generate just 200 kilowatts of power at peak output, enough to power perhaps 200 houses. But if everything goes according to plan, the farm could grow to 300 turbines.

The biggest tidal-power project to date is a



barrage at La Rance, France, with a capacity of 240 megawatts. Such barrages work like hydroelectric dams, holding back a head of water to power generators. But they are expensive and can damage estuarine ecosystems.

samples are tested further to gauge how much the virus has mutated. But as *Nature* went to press, WHO officials were still trying to arrange a meeting with a government minister, whose permission is needed before shipping samples overseas.

Biosafety lab gets a rocky ride in Montana

San Diego No one, it seems, wants the war on bioterror conducted in their back yard. On 12 August, three environmental groups sued the US National Institutes of Health (NIH) in an attempt to halt construction of a top-level biosafety laboratory in Montana.

The biosafety level 4 laboratory project at the NIH's Rocky Mountain Laboratories outside Missoula started about two years ago as part of a nationwide plan by the National Institute of Allergy and Infectious Diseases (NIAID) to boost research on pathogens that could be used in terrorism.

Plans for a similar NIAID-funded lab at Boston University are already meeting spirited resistance from local residents. Federal officials had hoped for an easier ride in rural Montana. But the environmental groups — Friends of the Bitterroot, Women's Voices for the Earth and the Coalition for a Safe Lab — have filed a suit alleging that the plan has proceeded without an adequate environmental-impact assessment.

Economist named as European science boss

Munich Scientists in Eastern Europe are celebrating the appointment of Janez Potocnik, from Slovenia, to the post of European commissioner for research.

The 46-year-old economist will succeed Philippe Busquin, a Belgian physicist who has been responsible for the research budget

of the European Union (EU) since 1999.

José Manuel Barroso, the designated president of the European Commission, announced his chosen team last week. The 25 commissioners make up the EU's executive.

Scientists in the ten states that joined the EU in May, who are worried about losing out in the scramble for EU funding, are delighted with Barroso's team, due to take office on 1 November. Hungarian László Kovacs will be responsible for energy, Cypriot Markos Kyprianou for health, and Slovak Ján Figel for education and training.

"This is a very promising signal," says Jerzy Langer, a solid-state physicist at the Polish Academy of Science's Institute of Physics in Warsaw, and an honorary member of Euroscience, a grassroots organization of European scientists.

NIH told to get tough on outside interests

Washington The National Institutes of Health (NIH) must draft tougher agency-wide regulations on conflicts of interest, says the Office of Government Ethics (OGE), which serves as a watchdog for the executive branch of the US government.

Responding to allegations made by the *Los Angeles Times* about lucrative outside consultancies held by NIH scientists, NIH director Elias Zerhouni proposed rules on 22 June that would prohibit paid consulting for senior employees. But the OGE suggests an absolute ban for all personnel. It also wants responsibility for policing the rules to lie with the NIH's parent body, the Department of Health and Human Services.

NIH deputy-director Raynard Kington says: "We will work closely with colleagues in the Office of Government Ethics to develop the new regulations for the agency. We are working hard to get this right."

The show goes on

CERN, the centre for particle physics in Europe, has been smashing its way through the subatomic world for the past 50 years. Alison Abbott finds out what's in store for the future.

During CERN's routine maintenance shut-down in 1999, acrobats and mime artists turned one of the vast halls within the laboratory's 27-kilometre tunnel into a stage. Light and shadows played around the nine-storey cavern, illuminating a poetic ballet of aerial gymnastics. On ropes, trapezes and ledges, the performers enacted Paul Dirac's creative struggle with his 1920s theory of antimatter and his pessimistic belief that such theories could never be confirmed experimentally. In the show's final scene, the huge backdrop parted to reveal the DELPHI experiment, where for the past decade CERN physicists have been observing collisions between electrons and positrons, their antimatter counterparts.

The show symbolized everything that CERN has achieved in its first 50 years (see '50 years of CERN', below). The lab's *raison d'être* has been to develop increasingly sophisticated machines to provide answers to the questions that plague theorists seeking fundamental truths about the Universe.

If science ever comes close to poetry, it does here, in the ugly sprawl of CERN's flat-roofed buildings overlooked by the beautiful Jura mountains near Geneva. Not just in the



Where art and science meet: the DELPHI experiment takes a starring role in a dance about antimatter.

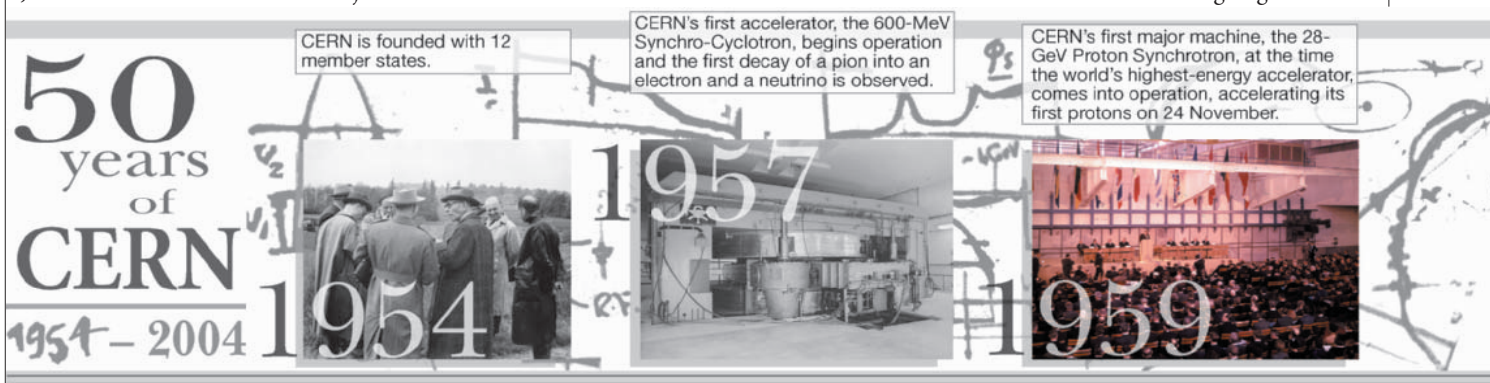
deep and fundamental questions with which the laboratory concerns itself — What is our Universe made of? How did it arise? — but also in the idealism of its scientific approach, and the charm of its working atmosphere.

Founded in the ruins of post-war Europe as a cross-border partnership to recreate the base in particle physics destroyed by the Second World War, CERN continues to live by its philosophy that science has no borders.

Ask any of the tens of thousands of physicists from 80 countries who have worked at

the lab over the past half-century, and exchanged views in its famous cafeteria, and they'll wistfully recall a touch of magic. "The intense but non-aggressive scientific style, the extraordinary variety of international scientists — I loved it, and it completely changed my perspective," says Harry Nelson, a particle astrophysicist at the University of California, Santa Barbara, who spent three years at CERN in the late 1980s as a postdoc. "And it's a place where some really great discoveries have been made. I sometimes dream of going back."

(FROM LEFT) CERN/SPL; CERN



But will the accolades still be coming in when CERN celebrates its 75th anniversary? The technical demands of particle physics have increased at an alarming rate — much to the dismay of governments that have to pay for machines that now cost billions. As purse-strings tighten, stresses have emerged in the traditionally cohesive international particle-physics community.

The community is expected to announce its technology choice for the next big machine by the end of this week. Despite the display of unity, behind the scenes CERN is gambling that an alternative technology it is developing will, at some future date, become a scientifically superior option.

Early days

For its opening act, in the early 1950s, the mood was upbeat. The concept of CERN was first proposed by the French physicist and Nobel prizewinner Louis de Broglie at the 1949 European Cultural Conference in Lausanne. Five years later, 12 states signed the founding treaty, making CERN the first post-war organization to include West Germany as a member. CERN now has 20 member states, and although the Soviet Union never joined, many of its scientists worked there alongside Westerners during the cold war.

Back in 1954, the frontier of knowledge was nuclear, not particle, physics, as CERN's full name, the Conseil Européen pour la Recherche Nucléaire, reflects. Theorists recognized three fundamental forces governing atomic matter — the strong nuclear interaction, the weak nuclear interaction and the electromagnetic force — but had little idea of how to integrate them in a way that would explain the physical Universe.

But they had some idea about what was missing. Although only protons, neutrons and electrons had been described as the basic constituents of an atom, cosmic-ray research had shown that there were a lot more exotic particles, with names such as 'pions' and 'muons', to be discovered.

The best way to make discoveries was to smash a beam of charged particles, such as protons or electrons, into a target at high speed



CERN's director-general, Robert Aymar, is responsible for ensuring the success of the lab's next big machine, the Large Hadron Collider, which will hunt the Higgs boson (simulated, right).

and then capture the resulting fireworks with cleverly designed detectors. The properties of newly discovered particles fed back into theories that became collectively known as the standard model of the structure of matter.

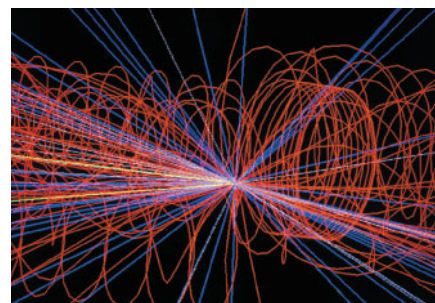
CERN's first machine, the Synchro-Cyclotron, accelerated a beam of protons up to energies of 600 million electron volts (eV) and smashed it into a fixed target. This relatively low collision energy was sufficient to generate pions, unstable particles formed from a mixture of matter and antimatter. The CERN pioneers were thrilled to watch pions decay into electrons and neutrinos in line with predictions from the standard model. Each new machine took collision energies higher, revealing new particles with properties predicted by the standard model.

CERN's second act opened in the early 1980s with technological innovation. In the late 1960s, CERN accelerator physicist Simon van der Meer had developed a key technology, called stochastic cooling, for storing antimatter particles. This made it possible to collide a beam of protons precisely with a beam

of anti-protons travelling in the opposite direction. If both beams are accelerated to similar speeds, the collision energy is doubled.

In the 1970s, the experimental physicist Carlo Rubbia, who held joint positions at CERN and Harvard University, championed this high-stakes technology. And in 1981, CERN upgraded its Super Proton Synchrotron (SPS) into a collider able to reach energies of up to 500 billion eV.

With its souped-up SPS, CERN scored two



great discoveries, the heavy W and Z particles, carriers of the weak nuclear force, just five months apart in 1983. This established CERN as a world leader, and won Rubbia and van der Meer the 1984 Nobel prize in physics (see 'Winning personalities', overleaf).

Upping the ante

CERN's most recent machine, the Large Electron Positron (LEP) collider, whose 27-kilometre tunnel straddles the Franco-Swiss border, hurled a beam of electrons into a beam of positrons. Unlike protons, electrons have no internal structure, so their collisions are less messy, although their lighter mass means that they cannot reach such high collision energies. Still, LEP was able to better characterize particles that had been discovered with the SPS, and was only switched off in 2000 to make room in the tunnel for its successor, the Large Hadron Collider (LHC). Until the LHC starts operating in July 2007, only a handful of small experiments are running at the lab.

As the machines get bigger, so do the collaborations. Rubbia's W and Z team involved 150 scientists. Today, teams at accelerator labs often number 500. The biggest experiment planned for the LHC, called ATLAS, involves 2,000 physicists.

Although the numbers may sound frightening, those working in big collaborations say that individuals are not condemned to anonymity. Michel Spiro, now director of the National Institute of Nuclear and Particle Physics in Paris, remembers working on the W and Z experiment: "Rubbia was number one, of course, but I made my own contribution which was recognized by my

"The non-aggressive scientific style, the variety of international scientists. I loved CERN, and it completely changed my perspective."

— Harry Nelson

Georges Charpak invents the multiwire proportional chamber that will win him the Nobel prize in 1992.



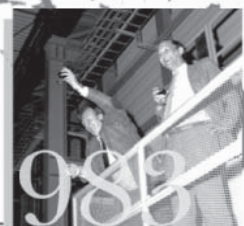
First observation of a neutral current, in which the charges of interacting particles — electrons or neutrinos — do not change.



The world's first proton-anti-proton collider, the Super Proton Synchrotron (SPS), comes into operation.



The W and Z bosons are discovered, which wins Carlo Rubbia and Simon van der Meer the 1984 Nobel prize.



colleagues.” And it all happened “in an atmosphere of thrilling enthusiasm”, he adds, “because we knew that the experiment could be a breakthrough for particle physics”.

In the ATLAS collaboration, experiments are split into subprojects. Karlheinz Meier of the University of Heidelberg, a principal investigator in ATLAS, says his own ‘Trigger’ project, which will design an electronics system to filter out the most interesting physics from LHC collisions, involves only 30 physicists from five institutes. “It’s sort of like the old days,” he says. “The group is small and we are getting to know each other very well.”

Project teams are scattered geographically, but share information over the World Wide Web, invented at CERN for this purpose. In future they will use the Grid, a global supercomputing network that CERN is developing to handle and distribute the vast data sets that the LHC will spew out.

Universal truths

The next half-century will be dedicated to research into the big unknowns. To complete the standard model, physicists predict the existence of one more particle, called the Higgs boson, which is needed to explain why particles have mass. Other theories that hope to extend the standard model predict ‘supersymmetrical’ partners of the known particles at higher masses. These particles have not yet been seen, but they should be created by the high-energy collisions, accelerating protons up to 7 trillion eV, within the LHC.

Discovering particles is one way to answer deeper questions about the Universe. We now know that the standard model describes barely 5% of the Universe. The rest is invisible to us. Some of this is thought to be composed of ‘dark matter’, which is apparent only through its gravitational pull and which might include supersymmetrical particles. This was joined in 1998 by the even more mysterious ‘dark energy’, whose existence was needed to explain observations that the rate at which the Universe is expanding is accelerating.

Theoretical physicists are already working on ideas that go beyond the standard model. “We have worked on overarching

theories such as ‘superstrings’ and ‘extra dimensions’ to try to explain how the Universe may fit together, but they are very speculative,” says John March-Russell, a theoretical physicist at the University of Oxford, UK, who spent four years at CERN. Like his colleagues, he hopes that the LHC will help reduce speculation.

“Knowing more about the Higgs and supersymmetry will let us pose the next questions more sensibly,” says Frank Wilczek, a theoretical physicist at the Massachusetts Institute of Technology.

As CERN reaches this dramatic crunch point — if the LHC does not find the Higgs, particle physics will be thrown into crisis — it takes centre stage in particle physics. No other organization has won political support for such a massive, yet esoteric, project.

The first sign that particle physicists had priced themselves out of the market came in 1993, when the US Congress pulled the plug on the planned Superconducting Supercollider, also designed to chase the Higgs. US physicists were glad that CERN member states held faith, because it gave them the opportunity to keep working. And when the LHC’s finances looked wobbly in 2001, the United States, Japan and Russia pitched in with in-kind contributions amounting to 12% of the overall building costs of SF\$3.2 billion (US\$2.5 billion).

CERN still pays the lion’s share towards the LHC’s infrastructure, but the detectors that will record the collisions are being built by 500 research institutes around the world from their own budgets. These institutes also provide most of the 5,000 or so physicists who will conduct the experiments, only two-thirds of

whom come from CERN member states.

This international mix will continue CERN’s ideal of science without borders for at least another two decades. “We have Indians working on the same LHC experiments with Pakistanis, Chinese with Taiwanese, Israelis with Arabs,” says CERN’s current director-general, Robert Aymar. “This ‘romance’ of CERN, the way it contributes to good relations, is not fading — quite the opposite.”

Aymar’s reputation is built on his management skills for large projects; most recently he headed the international ITER project to build a prototype fusion reactor. He is a pragmatist, who was brought in by the CERN council to restore political confidence in the LHC. Since his arrival in January he has restructured top management at the lab to ensure an iron grip on LHC development and costs. The collider’s finances are now assured, thanks to a system of bank loans that



Brain food: cafe society is at the heart of CERN’s relaxed atmosphere.



Surf’s up: the World Wide Web was devised at CERN.

“Knowing more about the Higgs and supersymmetry will let us pose the next questions more sensibly.”

— Frank Wilczek



must be paid back by the end of the decade. The world's most expensive ground-based science kit "will stay on target and within budget", Aymar promises.

"CERN will only be free to get involved in other projects after 2011, when the LHC debts are paid off," he adds. This is a crucial point, because CERN must think about life beyond the LHC, which is expected to operate until 2020, and possibly to 2025. The global physics community has already decided what machine it wants next, an electron-positron collider that will be able to characterize more precisely the particles that the LHC discovers. The machine of choice is a linear collider that can reach higher energies with electrons than did the circular LEP.

Two designs for such a sub-TeV linear collider, which will achieve energies just below 1 trillion eV, are on the table. The 'cold' design, which depends on superconducting technology, is led by DESY in Hamburg, and the 'warm' option by US and Japanese labs. The International Technology Recommendation Panel of the International Linear Collider Steering Committee met in Korea on 11–13 August to select one of these options for development by a global collaboration, with hoped-for operation in 2015. As *Nature* went to press, it was expected to announce its decision this week at the International Conference on High Energy Physics in Beijing.

This timetable would seem to rule out CERN as a host for the proposed machine, but with Aymar's arrival the stakes have been raised. Since 1986, CERN has been developing an alternative linear collider design called the Compact Linear Collider (CLIC). In a standard collider, the electromagnetic waves that accelerate the beams are created within special cavities called klystrons. Instead, CLIC uses a second set of intense low-energy beams to provide the waves that the accelerating electrons surf. This allows it, in theory at least, to achieve much greater accelerations. It may even be possible to achieve energies of between 3 trillion and 4 trillion eV.

Most physicists agree that, all things being equal, CLIC makes more sense scientifically. If the Higgs boson turns out to be heavier than expected, a sub-TeV machine may not be powerful enough to analyse it.

Winning personalities

CERN garnered two Nobel prizes in physics in its first 50 years, won by three very dissimilar characters. Simon van der Meer, from the Netherlands, and Carlo Rubbia, an Italian, shared the 1984 prize for the discovery of the *W* and *Z* bosons. Van der Meer was as shy as he was brilliant and inventive, whereas Rubbia's acknowledged genius was allied to a powerful personality.

Van der Meer shunned all publicity, whereas Rubbia became director-general of CERN from 1989 to 1994. His excitable nature generated some tension, but was offset by his impulsive genius, say some former colleagues.

The *W* and *Z* discoveries, and many others, would not have been possible without an earlier CERN invention for which Georges Charpak won the 1992 prize. Charpak's 'multiwire proportional chamber' replaced photographic methods for particle detection with electronic ones, improving both sensitivity and spatial resolution.

Charpak, born in Poland and now a French citizen, spent some of his earlier life in the French resistance and in the Dachau concentration camp. Elegant and charming, Charpak was reputedly ham-fisted in the lab and gave a handsome portion of his winnings to the technician who had helped him out.

A.A.

A higher energy range also means greater chances of spotting more supersymmetrical particles. And if the LHC sees no hint of the Higgs or supersymmetry, then the case for a sub-TeV machine looks distinctly weak.

But all things are not equal. The CLIC technology is far less advanced than the other designs. Until Aymar arrived, the project was progressing in a leisurely fashion towards a feasibility decision in 2013. He has brought this forward to 2009.

Future plans

This date is crucial. By then the LHC could have glimpsed the Higgs boson, and CERN will be almost free of its debts. The design phase of the sub-TeV machine will also be advanced, but it is unlikely that any country will have committed itself to the project's biggest expense — its construction. If feasibility studies show CLIC to be viable, it could present itself as an alternative. Spiro, who now heads the French delegation to CERN, is a strong believer in this strategy. "It would be better than committing to a less ambitious machine," he says.

But it's a long shot. Even if CLIC were shown to be feasible, it could not start operating until 2021 at the earliest, a long time for even particle physicists to wait. And the idea of building a sub-TeV machine first and CLIC later is not politically viable.

Although Aymar stresses that CERN's decision to accelerate CLIC should not influence plans for the sub-TeV collider, and he has never suggested that a CLIC machine,

if built, should be sited in Geneva, CLIC's new lease of life has ruffled the feathers of many of those involved in developing collider technology.

Part of the discontent is down to a perception outside Europe that CERN wants to consolidate its present leading position, even if this is at the cost of the well-being of the US particle-physics community, which, most agree, will suffer if all the big machines continue to be located elsewhere.

On balance, CERN's chances of hosting the next big accelerator are slim, so when it celebrates its 75th anniversary in 2029, it is likely to be a very different place. The stream of visiting scientists will have reduced to a trickle and the exciting discoveries will be taking place elsewhere. Nevertheless, Aymar is positioning CERN for a leading role in any future machine, wherever it might be located. He says he wants any European participation to be funnelled through CERN, rather than through national agencies or governments.

Even if CERN were no longer in the forefront, could its unique spirit relocate to the lab hosting the next big accelerator? Probably, although some worry that if current problems with visas continue, this would hamper any international project located in the United States. In the meantime, the place where the World Wide Web was born, and where fresh insights into the Universe will be argued over in that most celebrated of cafeterias, is set for a glorious finale.

Alison Abbott is *Nature's* senior European correspondent.

CP violation, the effect that helps to explain nature's preference for matter over antimatter, is confirmed by the fixed-target experiment NA48.



Thousands of anti-hydrogen atoms are produced and measured for the first time.



CERN celebrates its 50th anniversary, while construction of the LHC continues apace.



Pyramid power

Archaeologists have failed to learn the secrets of Mexico's largest ancient monument. Particle physicists might save the day, says Michael Hopkin.

It's not the kind of place you'd expect to find a particle-physics laboratory. The ancient Pyramid of the Sun at Teotihuacán is an hour's drive from Mexico City at the end of a bumpy jeep track. But deep inside this massive monument a state-of-the-art particle detector is being assembled, and my companions, both physicists, have brought me here to see how it's getting on.

The project is the brainchild of researchers from the National Autonomous University of Mexico (UNAM) in Mexico City, who hope to succeed where traditional archaeology has failed. Instead of taking up pickaxes and shovels to get at the pyramid's secrets, they will use their machine to detect the cosmic rays that continually pass through this mass of stone and dirt. With patience, the researchers believe, the rays will generate a picture of what is inside the monolith.

"The mass of the pyramid is so huge that we need twenty-first-century technology to study it," concedes UNAM archaeologist Linda Manzanilla, who is collaborating with the physicists.

The Pyramid of the Sun is believed to be the third largest pyramid in the world, with a base that is roughly 200 metres on each side. Built some 2,000 years ago, it formed the centrepiece of a bustling city for five centuries until Teotihuacán was abandoned. When archaeologists first dug into it in 1922, they hoped to find the bones of Teotihuacán's rulers. But unlike the neighbouring Pyramid of the Moon, which contains numerous stone chambers, the Sun pyramid is filled with earth and volcanic debris. Further digging seemed likely only to damage the structure.

Going underground

The physicists didn't need to dig. A cramped tunnel just below ground level that leads to several chambers at the heart of the pyramid was discovered in 1971. Although it provided scant clues for archaeologists as to what lies above, it made a perfect, if humid, spot for a particle detector.

As tourists admire the views from the top of the pyramid some 65 metres above us, we don hard hats and open a rusting metal door at the base of the great structure. My guides, UNAM particle physicists Arturo Menchaca and Ernesto Belmont, lead me down an iron staircase beyond the door to the tunnel opening. After several minutes of walking stooped through the stifling humidity of the tunnel we come to the metal shed that the researchers have set up to house their instrument.

Menchaca's team has been working on the



Catching rays: Arturo Menchaca (right) and Ernesto Belmont hope their muon detector will let them see inside the Pyramid of the Sun (above).

But he warns that even if something that looks like a room is spotted, it could simply be the result of subsidence over the centuries, rather than architectural design. "Everyone's expecting Tutankhamun," he says, "but we may find something that means nothing."

Subterranean blues

Doing science down here creates unusual problems. The researchers had to install their own power supply and encase it in a metal pipe to protect it from thieves. Once the detector is running, they will have to ventilate the chamber and tunnel before entering to avoid suffocating on the mix of carbon dioxide and noble gases used in the instrument.

No one should expect results any time soon. The researchers are testing parts of the detector in the lab and installing them underground piece by piece, a process they hope to complete within the next two months. They will then collect data for at least a year before releasing it to archaeologists. Should any of the tiny wires break, as has been known to happen, everything gets set back a week.

Perhaps there's no rush. As we emerge blinking into the sunshine to the curious stares of camera-clutching sightseers, the baking sun reminds me that it's better not to hurry in this climate. "This problem has been waiting 2,000 years," Menchaca says. "Nothing's going to happen if we're delayed another month."

Michael Hopkin is a reporter for news@nature.com.

Communication is key to aid development efforts

Government could hone its use of science but scientists need to understand the issues.

Sir—Recent calls by the United Nations (*Nature* **430**, 5; 2004) for stronger science input to support aid policy, in particular for feeding the hungry, are welcome. In the United Kingdom, organizations such as the Department for International Development (DFID) need to improve their use of the science base. But there is also scope for the scientific community to improve its understanding of development issues surrounding agricultural policy, if scientists are to be productively engaged in fighting world hunger and poverty.

In the United Kingdom, the call for better use of science in development has been led by the Royal Society and the science research councils. In a recent News story (*Nature* **429**, 492; 2004), John Lawton, of the Natural Environment Research Council, described DFID as “complacent, rather arrogant and ill-informed” about science.

However, similar shortcomings with respect to understanding of development issues are evident in the public pronouncements of some of the country's leading scientists.

At the same parliamentary inquiry at which Lawton addressed DFID's

shortcomings, John Pickett, of the Biotechnology and Biological Sciences Research Council, described how, on a visit to Malawi, his team was “whisked off” to view “some kind of DFID programme in which very, very small bags of seed and very, very small bags of fertilizer were being given out. ... This seemed to be a totally unsustainable and non-scientific based [sic] piece of development work which you would not really expect of an organisation like DFID”.

The programme Pickett refers to is known as the Malawi Starter Pack Programme, which, in the 1998 and 1999 planting seasons, aimed to supply Malawi's 2.8 million smallholder farming households with sufficient inorganic fertilizer and hybrid maize seed to plant 0.1 hectare (the average land-holding in southern Malawi is 0.3 ha). These “very, very small” inputs were intended to provide a short-term safety net, to enable Malawi's farmers to survive the consequences of the International Monetary Fund's Structural Adjustment Programme. This had withdrawn subsidies from agricultural inputs (including fertilizer), ordered a dramatic currency devaluation and caused (through

withdrawal of state services) the collapse of the agricultural credit system. As a result, most farmers were unable to afford the inputs needed to grow enough food for household consumption (see <http://web.africa.ufl.edu/asq/v6/v6i1a8.htm>).

Far from being unscientific, the Starter Pack programme was based on a thorough knowledge of the constraints faced by farmers and the production dynamics of Malawian agriculture. The programme was designed by Charles Mann, an economist at Harvard's John F. Kennedy School of Government and former food security adviser to Malawi's government. DFID provided much of the funding for the programme, but its implementation was a multi-donor effort.

With the secretary of state for international development, Hilary Benn, having recently confirmed that DFID will appoint a chief scientist, there are good opportunities for science to serve development needs, provided there is effective communication on both sides.

Edward H. Allison

*School of Development Studies,
University of East Anglia,
Norwich NR4 7TJ, UK*

Proud past but no future for pioneering institute

Sir—As reported in your News story “Institute doomed by loss of interest in basics” (*Nature* **430**, 282; 2004), Japan's Biomolecular Engineering Research Institute (BERI) will disband sometime next year. Unfortunately, for various reasons, including the fact that the board of directors had not yet met to finalize its decision, I was unable to comment at the time.

During the past 18 years, BERI has achieved much in the field of protein structure and functional analysis. Many scientists starting at BERI have gone on to excellent positions in industry or academia. BERI has been highly reputed both at home and abroad and it has amply succeeded in attaining its original goals.

But in changing times that increasingly call for research geared towards practical developments, it has become difficult to reconcile the interests of basic research with varied corporate needs.

BERI's project grants from the government have fallen by half during recent years, increasing the burden on

industry. And a government evaluation on 25 March this year, while praising the projects' technological merits, criticized their wider applicability.

As chairman of the board of directors and a founding member of BERI, I find it regrettable but unavoidable that the nine companies supporting BERI have decided, in consultation with the Ministry of Economy, Trade and Industry, to disband it.

Many important issues remain, such as how to bring ongoing projects to fruition and how to support the relocation of experienced researchers and staff. I predict that there will be many employers eager to take them in.

Tadashi Hirata

*Biomolecular Engineering Research Institute,
6-2-3 Furuedai, Suita-shi, Osaka 565-0874, Japan*

Plant biologists need to get back to their roots

Sir—Sean Nee is right in drawing attention to the invisible world of microbial life in his Commentary “More than meets the eye” (*Nature* **429**, 804–805;

2004). That netherworld of life is not the only one to suffer neglect on the part of biologists, however. The microscopic life in the soil depends in large measure on plants exporting to their roots the sugars produced by photosynthesis in their leaves. But biology—even plant biology—is obsessed with the visible plant.

Mother Nature conspires with biologists to keep plant roots in the dark. Yet all visible life on the surface depends, for most of the chemical elements it requires (the mineral nutrients), on plant roots which are ‘out of sight, out of mind’.

Anyone wanting to test the neglect of this matter need only try looking up ‘plant nutrition’ in university catalogues. That science, so prominent half a century ago, has all but lost an identity of its own. Pretending that this subject belongs in soil science is like considering photosynthesis as part of atmospheric science. The roots of life on Earth deserve better.

Emanuel Epstein

*College of Agricultural and Environmental Sciences,
University of California, Davis,
One Shields Avenue, Davis,
California 95616, USA*

Raising Europe's game

How to create a research council that is a Champions League for science.

Robert M. May

The first pan-European meeting on science and technology in society is being held this month in Stockholm¹. As you might expect, or perhaps dread, there is a session on the proposed European Research Council (ERC), now being discussed by the European Commission in Brussels. Although Britain's Royal Society and I are enthusiastic supporters of this idea^{2,3}, we feel a new council must not be hamstrung by unrealistic expectations, whether those consist of closing the overall research and development funding gap with the United States or of capacity building in the ten new members of the European Union (EU).

My personal enthusiasm for the ERC is based on spending the past 15 years in Europe, and having in my own research group people from 12 different countries and five continents, some of whom were supported by European postdoctoral programmes. These people are testimony to the strength of Europe and to the fact that we already have some good programmes for launching the careers of young scientists.

But if we are to implement the wider vision of a common European Research Area, we should aim to create an ERC that adopts the virtues of the US National Science Foundation. For this to work, the ERC must be based purely on principles of peer-reviewed scientific excellence; it must have minimum bureaucracy; and it must not exist at the expense of national funding.

We need to strip away some of the rhetoric that clouded earlier discussions. Many advocates of an ERC lament the gap in research and development spending, as a percentage of national wealth, between the United States and the established EU members (EU15)^{2,3}. But look more closely at the data^{2,4} (see top graph overleaf) and it is clear that almost all (96%) of the EU15 funding shortfall is due to lack of spending by business and industry, which focus more on development than research. In this context, the Lisbon and Barcelona declarations of 2000 and 2002, with their aim of surpassing the United States in research and development spending by 2010, seem astonishingly naive. I do not believe there is a snowball's chance in hell of this happening, but whether or not this improbable ambition is fulfilled lies entirely in the hands of industry, not government.

Look at the total numbers of researchers in universities and government laboratories in the EU15 compared with the United States



Flagging up issues: lack of bureaucracy and realistic goals will be vital in a European Research Council.

and again the pattern is clear^{2,4} (see bottom graph overleaf). The difference between the United States and the EU15 is in researchers employed by business and industry, not the science base.

Basic skills

Why do governments spend taxpayers' money on the science base? Why did the British Prime Minister Tony Blair say⁵ "the science base is the bedrock of economic performance"? It is for the new knowledge produced. More importantly, it is for the cadres of young people who emerge, some

cycling back into knowledge production, others carrying the knowledge out into business, industry and elsewhere.

There is essentially no difference between EU15 and US spending on basic science^{2,4,6,7}. As discussed by David King⁸, Britain's chief scientific adviser, these statistics conceal countries within Europe, such as Sweden or the Netherlands, that proportionately spend significantly more and produce more in papers, or citations, than does the United States. Other countries spend more (France and Germany, say) and get less in output than the United States, and others spend less

F. LENOUR/REUTERS

(for example, Britain) and produce more, in relation to population size or the amount spent. And there are, sadly, several countries within the EU15 that both spend less, proportionately, and get less than the United States. How you spend the money is at least as important as how much you spend⁶.

Against this factual background, I offer some personal opinions about what an ERC should be and do. My focus is on basic research, because I see this as the motive for, and the purpose of, an ERC. How to foster the research and development that is devoted to innovation and the adventure that carries knowledge into the marketplace is a separate issue (albeit one where the EU15 greatly lags behind the United States).

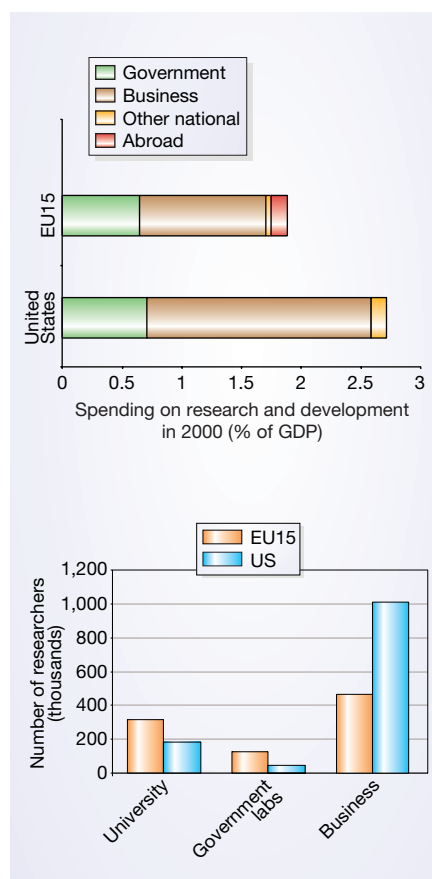
Golden goals

A primary reason for having an ERC is that it will create a larger playing field, helping to raise aspirations. Because different countries have different scientific cultures and ways of doing things, some of which are demonstrably more productive than others, it makes sense to create a mechanism that could raise the standards of European science in the same way that the Champions League improves national football leagues.

There are at least five other motivations for an ERC, which I offer here. The first two I regard as absolutely essential. I do not think any of the Nobel laureates who signed the long document of support for the ERC would be happy were these first two constraints not satisfied. The others we could, I think, argue about.

The first condition has been put forward many times; there is, however, a difference between talking the talk and walking the walk. Everyone agrees that the ERC must be built on peer-reviewed, uncompromising excellence. What this means in operational terms is that the director has to be a scientist of international stature: someone who knows what excellence and peer review really mean. It cannot be a Eurocrat or an apparatchik, no matter how eminent. The council should not be too large: a dozen or so people representing the worlds of science, innovation, business and industry — all of whom are still active in those worlds. The council should be chosen by soliciting advice from the constituent national academies or the European Academies' Science Advisory Council.

Much of the rhetoric about the ERC makes comparisons with the United States when these are convenient, so let us compare the structure of our proposed ERC with the National Science Foundation. The foundation has a small council, which does not take a representative from each of the 50 states. Naturally, it is subject to worries that, for example, Oklahoma does not get as much money as California. But the council members are appointed simply as unambiguous examples of excellence in basic science.



This is what the ERC needs.

A second constraint concerns the urgent need to build scientific capacity in some of the new members of the EU25, and arguably even in some of the EU15. Whereas it is vital that this issue be constructively addressed, it is equally vital that it be seen as a separate issue from the creation of a peer-reviewed, excellence-based ERC. I think it is really the EU structural funds that should be used — alongside their current uses — for the hugely important endeavour of capacity building for the knowledge economy.

Studies suggest^{6,7,9} that some 50% of productivity growth in particular industries and, more generally, in certain countries, comes from new knowledge rather than labour or capital. This is a compelling argument for addressing scientific capacity building through EU structural funds. If we do not address these needs immediately, there will be intolerable pressures upon the ERC to serve two conflicting goals, one addressing excellence in scientific proposals, and the other creating the infrastructure for building future excellence.

Third, I greatly admire the way the European Science Foundation has created collegial networks, and yet I recognize that such networks struggle to find funding that crosses national boundaries. The Human Potential programme and the Marie Curie Actions programme bring the best young people together in the best laboratories in Europe. Many of them wish to go back

to their own countries, while continuing collaborations with their new colleagues. I think Europe needs a mechanism that evaluates grant proposals from groups spanning national boundaries using the pure criterion of excellence, just as the best national research councils do.

We should recognize the urgency of this need. In Britain, half the papers published in science, medicine and engineering involve collaboration between two or more different institutions. From bibliometric statistics⁸, it can be deduced that collaborations across national borders account for a large, and fast growing, fraction of all 'elite' publications (top 1% of citations) within the EU15: specifically, 22% of such papers in 1993–97, rising to 29% in 1997–2001. We need funding mechanisms that recognize these trends, and it seems the ERC would be ideal.

My fourth point focuses on helping the best young people to pursue their own ideas in the best laboratories, free from hierarchical constraints. Existing EU postdoctoral programmes already do this very well, but they could be even better. They suffer from excess bureaucracy, especially when compared with the best postdoctoral programmes in individual countries. Furthermore, the better national programmes have realistic 'dowries', so that the recipients are not obliged to appeal to local sponsors for help with travel and other 'extras'. I would like to see the existing EU15 schemes gradually transferred, or parallel schemes created, as part of the ERC.

My fifth observation is one with which I believe many would agree. But equally, it will excite some opposition. I believe it is important that the independence of the ERC be underlined, both for substantial and for symbolic reasons, by making sure its headquarters are outside Brussels.

In building our vision of One Europe we have got off to a better start than we often acknowledge, especially when drawing unflattering comparisons between Europe and the United States regarding basic research⁸. I recognize that there will be understandable opposition to some of the above ideas, but little of it will come from scientifically excellent institutions or individuals, and least of all from the aspiring younger scientists who are Europe's future.

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. www.esof2004.org
2. Future Funding of the European Science Base (The Royal Society, London, 2004); www.royalsoc.ac.uk/policy; ref: 17/04
3. The Place of Fundamental Research in the European Research Area: Response to the Mayor Report (The Royal Society, London, 2004); www.royalsoc.ac.uk/policy; ref: 08/04
4. OECD Basic Science and Technology Statistics (OECD, Paris, 2001).
5. Blair, T. *Science* **281**, 1141 (1998).
6. May, R. M. *Science* **275**, 793–796 (1997).
7. May, R. M. *Science* **281**, 48–51 (1998).
8. King, D. *Nature* **430**, 311–316 (2004).
9. Stiglitz, J. et al. *Supporting R&D to Promote Economic Growth* (Council of Economic Advisors, Washington DC, 1995).

Seeing stars in a big way

The Gemini project typifies the growth of astronomy into 'big science'.

Giant Telescopes: Astronomical Ambition and the Promise of Technology

by W. Patrick McCray

Harvard University Press, 2004. 376 pp.

\$45, £29.95, €41.50

Sidney C. Wolff

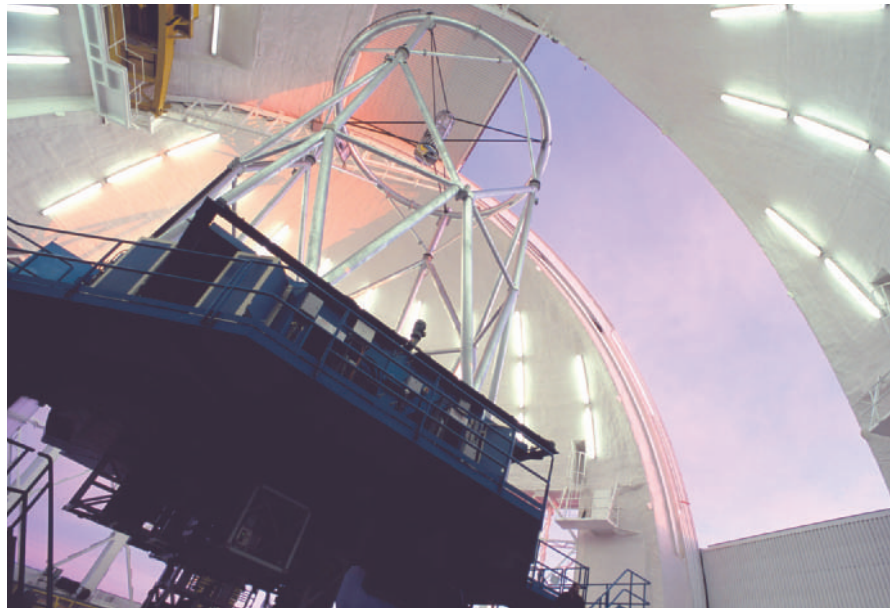
In the past 25 years, the number and size of ground-based optical and infrared telescopes has increased at an unprecedented rate. *Giant Telescopes* by science historian W. Patrick McCray describes part of this rise, the Gemini project, which led to the creation of two telescopes with mirrors that are 8 metres in diameter, the first in Hawaii and the second in Chile. McCray focuses on US involvement in the project, but Britain, Canada, Argentina, Brazil, Chile and Australia are also partners in Gemini.

The book begins with the completion in 1948 of the 5-metre telescope at Palomar Mountain, California. It then describes the controversy that surrounded — and still surrounds — the establishment and definition of the role of the US National Optical Astronomy Observatory (NOAO). Should it provide flagship facilities to support cutting-edge research? Or should it merely be, as many astronomers at elite universities have argued, one of several equivalent observatories with the responsibility of providing observing time to astronomers at institutions that do not have their own facilities? Even now, nearly 80% of telescope surface available to US astronomers is owned by university and private observatories, which restrict access primarily to their own staff.

Giant Telescopes discusses the technical breakthroughs in developing lightweight mirrors that enabled the construction of these telescopes. Three different technologies were developed in the 1980s: thin, flexible meniscus mirrors; stiffer, honeycombed borosilicate mirrors; and segmented mirrors. All of these have now been incorporated in telescopes that deliver excellent image quality. McCray summarizes some of the scientific questions that stimulated the worldwide investment in new telescopes.

The efforts of the NOAO to initiate a giant-telescope project are then described. After many false starts, the NOAO successfully nurtured the Gemini project and identified international partners. The Gemini telescope in Hawaii had first light in 1999, the one in Chile a year later, and both telescopes are now operating successfully.

McCray has relied on written documentation and interviews with the key players in the Gemini programme to tell his story. The



A clear vision: the Gemini South telescope in Chile is part of a large international project.

interviews were conducted while memories were still fresh, so the book is a valuable record of a critical period of astronomical development. Not only was a new generation of telescopes under construction, but astronomy was making the transition to becoming 'big science'. It was moving from an age when instruments and telescopes were built by small groups led by innovative astronomers who made real-time decisions about what to observe to a time of projects costing hundreds of millions of dollars that required large teams with specialized engineering, software and management expertise. McCray examines the social, fiscal and institutional forces that influenced postwar astronomy and other sciences that were dependent on ever-increasing financial investment.

As a participant in many of the events described, I found the book to be accurate overall. McCray does a good job of representing the various points of view about controversial issues without making value judgements, and his lack of bias makes this a particularly useful record. Equally importantly, the book is quite simply a good read. The explanations of both technical and scientific issues are clear and correct, and the book can be easily understood by anyone with an interest in either astronomy or the sociology and politics surrounding big-science projects.

US ground-based astronomy has historically been dependent on a pluralistic approach involving private philanthropy, universities and federal funding. Europe, in contrast, has chosen to unify its efforts

through the European Southern Observatory. Which will prove to be the best approach now that astronomers are initiating technical studies for still larger telescopes? Europe has set as its goal a 50–100-metre facility, whereas US astronomers are designing 20–30-metre telescopes. Are the more modest US aspirations a correct assessment of what is technically feasible or a sign that the fragmented US community has finally ceded leadership in ground-based astronomy to Europe?

There are two competing designs in the United States. One is based on compound, segmented mirrors, like those on the 10-metre Keck telescopes on Mauna Kea in Hawaii. The other would make use of seven separate mirrors, each 8.4 metres in diameter, on a single mount. This competition is eerily reminiscent of the shoot-out in the mid-1980s, described by McCray, between two approaches to building a 16-metre telescope for the US community. In the end, that telescope was never built, in part because of the failure of US astronomers to coalesce around a single project. There are many lessons in McCray's book for those who would avoid repeating history. ■

Sidney C. Wolff is at the National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, Arizona 85719, USA.

More on telescopes A Visitor's guide to the Kitt Peak Observatories

Leslie Sage & Gail Aschenbrenner
Cambridge University Press, 2004
\$15, £12.99

The course of true science

Investigative Pathways: Patterns and Stages in the Careers of Experimental Scientists

by Frederic Lawrence Holmes

Yale University Press: 2004. 288 pp. \$35, £25

Robert Olby

Recollections can be vivid, as when Francis Crick looked back to a moment in February 1953: "Jerry Donohue and Jim Watson were by the blackboard and I was by my desk, and we suddenly thought, 'Well, perhaps we could explain 1:1 ratios by pairing the bases.' It seemed too good to be true." Was this one of those 'eureka' moments, like August Kekulé's vision on the Clapham omnibus of a snake biting its tail, that led to propose the ring structure of benzene? Should we take these personal recollections for gospel, or seek out the written record, pour over the research notebooks for the embryo of the revelation, the stages in its almost subconscious formulation?

This and other questions concerning the career of the experimental scientist are raised and discussed by eminent science historian Larry Holmes in *Investigative Pathways*. Written shortly before his death, this brief book offers his reflections on 45 years of research into the careers of outstandingly successful experimental scientists, such as Claude Bernard, Antoine-Laurent Lavoisier and, most recently, Seymour Benzer. Holmes also looks back on his detailed study of the famous experiment by Matthew Meselson and Franklin Stahl on the semi-conservative replication of DNA. As well as mining his own intimate knowledge of these scientists, Holmes draws on the scholarly researches

of David Gooding on Michael Faraday, of Martin Rudwick on the geological 'Devonian controversy', of Gerald Geison on Louis Pasteur, and of Nicolas Rasmussen on the electron microscope.

This book is no mild valedictory. Rather, Holmes seeks to persuade us that his concept of the investigative pathway provides the framework within which to view the research careers of these scientists. The great experiments, discoveries and eureka moments do exist, but are the nodal points in the investigative pathway. This pathway has a continuity that survives the surprises that nature throws at us. Yes, the experimental system is an important element in the story. Sometimes it takes the lead, redirecting the researcher, but it rarely lifts him out of the pathway being investigated. Even the belated recognition that the ribosome is not the message did not throw research into protein synthesis off course, for example. But such events serve to warn us against teleological reconstructions of the past in which it is assumed that the end point finally reached was envisaged from the start.

Holmes' intimate knowledge of the research careers of his subjects is clearly apparent. This book serves admirably to introduce the reader to his many studies and those of his colleagues in the field. For the biographer he offers insights into such topics as mentoring, creativity, the difficulty of remaining at the forefront as a scientific field matures, and the problems of ageing for the eminent scientist. Holmes is no anthropologist bereft of scientific knowledge coming to the lab to report on the strange society within, but a scholar who has devoted his life to understanding what goes on there.

Fashions in historiography come and go but Holmes, while absorbing what is valuable in each, has remained true to the calling he felt from the beginning: to seek to under-

stand the stepwise generation of scientific concepts in the experimental life of the scientist. Departing in significant ways from other researchers, such as Thomas S. Kuhn, from the Edinburgh school of the sociology of knowledge and from the ethnomethodologists, Holmes found that the investigative pathway provided him with the best framework in which to place his detailed historical accounts. It expressed, he felt, the "distinctiveness and continuity of the individual scientific personality".

Comparing the metaphor of the path to Howard Gruber's 'network of enterprise' and Gooding's 'experimenter's space', Holmes admitted that his investigative pathway would be difficult to apply where the scientist engages in several research topics and moves back and forth from one to another, or leads a team of researchers. Confining his chosen cases to those in which his subjects worked alone and did not make such shifts, or to episodes in a scientist's life when he worked in this way, has enabled Holmes to exploit the pathway metaphor effectively.

Historians and scientists will find this little book both stimulating and informative. It will surely join that select group of classics that long outlive their authors. ■

Robert Olby is in the Department of the History and Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

A struggle for order

A Well-Ordered Thing: Dmitrii Mendeleev and the Shadow of the Periodic Table

by Michael D. Gordin

Basic Books: 2004. 336 pp. \$30, £22.50

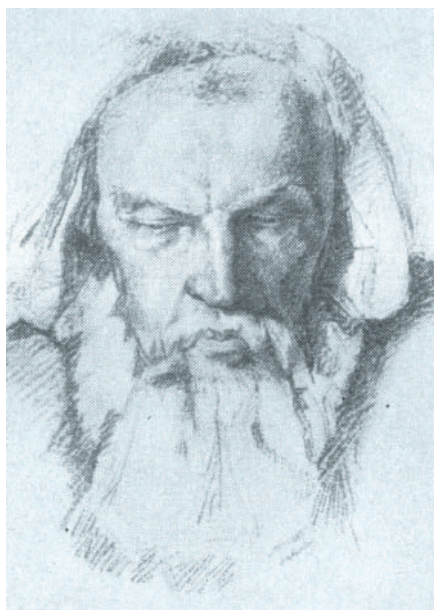
Bernadette Bensaude-Vincent

The name of Dmitrii Mendeleev is forever associated with the periodic table, which is found in chemistry laboratories and classrooms around the world. Yet this famous invention, which made sense and order out of the elements, was just one of Mendeleev's numerous achievements. Michael Gordin, an assistant professor of history at the University of Princeton, has reconstructed Mendeleev's heterogeneous career in all its facets and with all its contradictions. His book, *A Well-Ordered Thing*, is neither a standard scientific biography nor an attempt to demystify this scientist, who became a national icon in Russia. Rather, Gordin uses Mendeleev as an example to explore the life and work of members of the educated élite in the nineteenth century in imperial St Petersburg.

Historians of chemistry might well feel a bit frustrated because there is little chemistry in this book. The need for a means of teaching chemistry was crucial to the



Life in the lab: Claude Bernard (third from right) teaching, in a painting by Leon L'hermitte.



Dmitrii Mendeleev's periodic table was in tune with imperial Russia's desire for social order.

creation of the periodic system, so Gordin might have done well to give more attention to Mendeleev's textbooks. His 1861 organic-chemistry textbook is dealt with too hastily, with the excuse that it was quickly eclipsed by Aleksandr Butlerov's book. And Mendeleev's successful *Principles of Chemistry* could have been analysed against the background of the tradition of university textbooks. Gordin also provides no details about earlier attempts at classification, or about how the periodic system was received either in Russia or abroad.

But the history of chemistry is not Gordin's main focus. Instead he attempts to understand the cultural impact of the major reforms and political upheavals that occurred in imperial Russia before the end of the nineteenth century. From this perspective, Mendeleev's periodic system appears as a metaphor underlying a programme for restructuring and modernizing tsarist Russia. The periodic law, with its predictions of unknown elements and bold corrections of atomic weights, was one expression of an irrepressible attempt to eradicate misfits and anomalies in various domains.

This long-standing quest for order contrasts with Mendeleev's versatility. Although for today's chemists he embodies the chemistry of the elements, he did not spend much time on this topic. He abandoned his research on elements soon after constructing the periodic table, despite uncertainties about the classification of rare-earth elements and rare gases. In the 1870s he initiated a project that was his age's equivalent of 'big science' because it involved high-pressure devices. His objective was to investigate deviations from the ideal gas law with the expectation of isolating ether, an unknown, all-pervading substance that was postulated

by both Newton's dynamics and James Clerk Maxwell's electromagnetism. Mendeleev's ambition was to integrate ether as a chemical element within the periodic system, in order to unify the natural sciences. He also sought to save the individuality and integrity of chemical elements, which were threatened by radioactivity and electrons — the existence of subatomic particles favoured the view that atomic elements were made up of smaller units.

In the name of science, Mendeleev spent his life fighting against 'deviations' or superstitions. For example, he struggled against the fashion among educated people for spiritualism, and set up a commission for investigating mediums at the Russian Physical Society. Mendeleev was also concerned with the public face of science. In the newspapers and in his books, Mendeleev defended the legitimacy and the authority of scientific societies in matters of public opinion. He acted as an expert, first locally and then at the national level, notably through his work on standardization at the Bureau of Weights and Measures and in his attempt to modernize the calendar.

Gordin portrays Mendeleev as a loyal

subject of the Tsar, with conservative ideals, who fought desperately against the disintegration both of the Russian Empire and of chemical elements. He never really separated in his mind the future of Russia from the future of science, and had ambitions to be the Russian Newton.

This highly readable book offers two important lessons for working scientists. First, Mendeleev's career illustrates the interplay between scientific creation and economic, political and educational projects. Second, it may be a consolation to know that such a well known scientist endured an incredible number of failures throughout his life. Notably, his project to isolate ether failed and affected his scientific credibility. His solution theory and his views about the origin of oil were wrong. He also failed to reform the calendar, and his application to the Imperial Academy in St Petersburg was turned down. But above all, his firm belief in the individuality of chemical elements — the firm ground in which the periodic system was rooted — finally crumbled.

Bernadette Bensaude-Vincent is in the Department of Philosophy, Université de Paris X, 92001 Nanterre Cedex, France.

An architectural aside

The Italian scientific revolution, championed by Galileo in the seventeenth century, shares its roots with the mathematical beauty of Renaissance architecture. Galileo, for example, found that studies by the sixteenth-century master architects Giorgio Vasari and Michelangelo came in handy for computing the height of mountains on the Moon. And the Roman architect Marcus Vitruvius Pollio — whose *Ten Books on Architecture* (*De architectura*) is still required reading for architecture students today — drew on the proportions of the human body to scale his buildings, on astronomy for their accurate orientation with respect to the heavens, and, of course, on geometry.

The extensive margins of the first printed version (incunabulum) of the *Ten Books on Architecture*, in 1486, allowed for hand-written annotations. In 1520, Giovanni Battista Sangallo, a leading figure of sixteenth-century Roman architecture who worked with Raphael on St Peter's Basilica in Rome, filled the margin with annotations and beautiful drawings.

To celebrate its 400th anniversary last year, the Accademia Nazionale dei Lincei, Italy's national academy, has published a high-quality facsimile of this copy of *Ten Books on Architecture*. The page shown here describes the importance of the orientation of buildings for the health of their inhabitants. In a series of illustrated comments, Sangallo stresses the relevance of this for the Roman climate.

This exquisite book links the genius of

Vitruvius' original text to its first printed edition, to the freshness of Sangallo's notes, and to a contemporary introduction by art historian Ingrid Rowland.

Giovanni F. Bignami



The blind leading the sighted

An eye-opening experience of the wonders of perception.

Richard Gregory

One of the most dramatic, though very rare, human experiences is the acquisition of sight as an adult following blindness from birth. This happened to Sidney Bradford at the age of 52, having been recorded as blind at ten months and almost certainly blind, owing to an infection, from birth. Studying Bradford's new vision, and how it developed over the few years remaining to him, was the turning point at least for my appreciation of the wonders of perception.

It was my research assistant, Jean Wallace, who saw in a local paper in 1959 that a man, blind from birth and now 52, was about to receive new eyes — or rather, corneal grafts that might allow him to see. We gathered up every visual trick and gadget we thought might be relevant from the practical class we ran in the Cambridge Psychological Laboratory. Dropping teaching and other commitments, we drove to a Midlands hospital in time to see Bradford just after his first eye was opened to the light.

We found a cheerful, confident, middle-aged man who was willing to be investigated and who, so far as we could tell then or later, was truthful and honest. But an initial shock nearly made us turn back with the disappointment that this must be a put-up job, or at least a Great Mistake: he correctly read the time on the clock in the ward. Could he have guessed it? Borrowing a nurse's alarm clock, we set its hands to various positions, and he told us the times it showed. Taking a large watch, which had no glass, from the top pocket of his jacket, he told its time by rapidly touching its hands, as he had done for many years. So he could see immediately, from earlier touch experience. At least for us, this was a turning point for understanding vision.

Also rather by accident, we found he could read the name of the popular *Everybody's* magazine, guessing the title from its upper-case letters even though he could not read the lower-case ones. It turned out that at the blind school he had attended, the boys were taught to read upper-case letters by touch, from wooden blocks engraved for the purpose. Technically, Bradford showed cross-modal transfer from touch to vision, which no one knew about at that time, although it was soon discovered by other researchers studying primates.

The classical reference for considering the possibilities of adult sight given infant blindness is John Locke's *An Essay Concerning Human Understanding* (1690). Locke quotes a letter from his friend William Molyneux, posing the question: "Suppose a man born



Common senses: as Sidney Bradford embraces sight, his experience of touch informs his new faculty.

blind, and now adult, and taught by his touch to distinguish between a cube and a sphere... [Then suppose] the blind man be made to see: query, whether by his sight, before he touched them, could he distinguish and tell which was the globe and which the sphere? To which the acute and judicious proposer answers, 'Not. For, though he has obtained the experience of how a globe, how a cube, affects his touch, yet he has not yet obtained the experience, that what affects his touch so or so, must affect his sight so or so.'" The philosopher George Berkeley (1685–1753) held the same opinion. And so, essentially, did psychologists at the time Bradford gained his vision, notably Donald Hebb. This distinguished Canadian psychologist, author of *The Organization of Behaviour* (1949), coined the idea of neural nets and believed sight to be a skill acquired only gradually.

Early cases of operable blindness (reviewed by Marius von Senden in 1932) involved lens cataracts, sight being given by removal of the lens. But as these primitive operations did not allow a decent retinal image for weeks or months, patients' slowness to see was interpreted as evidence that they had no initial vision. It is now known that babies a few hours beyond birth take interest in faces. Their responses to stimuli important to survival, such as facial expressions, suggest considerable immediate vision, as Bradford had. But Hebb was right that visual learning is important and slow — it was only knowledge acquired from touch that enabled Bradford to interpret an artificial stimulus, such as a clock face, correctly.

Bradford's responses to well-known illusion figures were far from normal. He perceived far less distortion, and he did not experience the flipping ambiguity of the Necker Cube, or other such dynamic changes

of appearance. Pictures looked flat and meaningless. Perspective meant nothing to his visual system, yet he could judge the distances and sizes of objects that were already familiar from touch, such as chairs scattered around the ward — although he was wildly wrong about distances to the ground from the hospital windows. Evidently, earlier touch experience and behaviour such as walking, calibrated and gave sense to his vision, which was almost useless for untouchable objects or pictures. His unusual responses to the figures suggested that many illusions result from cognitive processing, rather than physiological signal processing occurring early in the visual system; this led to experiments and interesting controversies that persist today.

Vision was generally thought to be separate from the other senses, and is still mainly studied in isolation; yet Bradford showed that exploratory touch — as well no doubt as taste, sound and other sensory experiences — gives richness and meaning to retinal images. For optical images are but ghosts, materialized into objects by perceptual experience of the non-visual properties of things.

Now there is a new case. Mike May, in California, lost sight in both eyes from a chemical accident when he was three-and-a-half years old. He regained sight by a corneal graft, but the unique operation required the use of stem cells, which is why this was delayed until 2000, when he was in his early forties. May's first visual experiences, and his currently developing vision, mirror those of Bradford's late entry to the world of sight. And what is happening in his brain can now be investigated, thanks to another miraculous turning point: the advent of brain imaging technologies. ■

Richard Gregory is in the Department of Experimental Psychology, University of Bristol, Bristol BS8 1TN, UK.

Hitting the sweet spot

David A. Tirrell

By taking advantage of the cell's carbohydrate metabolism, reactive sugar analogues can be used to tag specific cells, potentially singling them out for imaging studies or drug delivery.

Delivering chemical agents to specific locations in the body has been an elusive goal ever since Paul Ehrlich described his concept of the 'magic bullet' nearly a century ago. Whether the objective is delivering drugs to sites of injury and disease, or the infusion of imaging agents that allow physicians to see what is going on, one would like to be able to target molecules to the places where they are needed most. Ideally, one would be able to pick out a single cell (such as a cancer cell) in the midst of many others, and deliver a molecular payload to that cell alone.

Given the subtlety of the differences between cells, it is not surprising that our bullets are somewhat less magical than we would like. To be sure, there have been successes in targeting using antibodies¹ and ligands for cell-surface receptors²; nevertheless, the specificity of such targeting methods remains modest — and problematic with regard to practical application.

On page 873 of this issue, Bertozzi and colleagues³ describe a new approach for differentiating one cell from another. First, mice are treated with a reactive sugar analogue that makes its way to cell surfaces by one of the normal pathways of carbohydrate metabolism. Cells that metabolize the analogue are thereby labelled with chemical groups that do not normally appear on their surface. Next, the mice are fed with a chemical agent that reacts specifically with the newly arrived sugar labels and leaves unmarked cells alone⁴ (Fig. 1).

The fact that this works is remarkable. Success requires that the sugar analogue finds its way to the cell surface through the appropriate biosynthetic pathway, that the reagent delivered in the second step is selective enough to react only with labelled cells, and that the reagent forms a new chemical bond *in situ*. Bertozzi and colleagues show convincingly that each of these requirements can be met; labelling is readily detected in mice treated with both reagents, but not in those treated with either reagent alone. There is also evidence that the level of labelling varies from one organ to another: in preliminary experiments, heart, liver and kidney were labelled more heavily than brain or thymus tissue, although it is not yet clear why this is so.

What might be done with this new approach? The most obvious opportunities are in fundamental studies of cell-surface

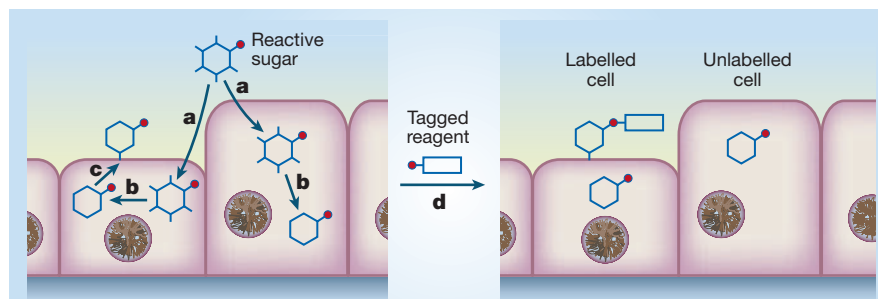


Figure 1 Chemical labelling of cell surfaces in live mice. Bertozzi and colleagues³ report an innovative approach to targeting chemical agents to specific locations in the body. First, the mice are treated with modified sugars, which are taken up by cells (a), undergo metabolic transformation (b) and are displayed on the cell surface (c). Next, a tagged reagent reacts specifically with the modified cell surfaces (d). Cells that do not display the modified sugars are not labelled.

sugars. The pattern of sugars on the cell surface is crucial in cell–cell communication, in interaction with pathogens, and in mediating inflammation and disease⁵. The labelling strategy described by Bertozzi and colleagues allows one to probe the set of sugars arrayed by the cell, to explore biosynthetic pathways and to examine the functional consequences of modifying the complement of cell-surface sugars. The authors suggest, for example, that the relatively high levels of labelling observed in heart and kidney might be a consequence of the fact that these organs are deficient in one of the enzymes involved in the synthesis of the natural sugar that competes with the labelling analogue.

This example also demonstrates how exceptional this approach to the labelling and targeting of cells is. Ordinarily, a cellular target is selected because it is a molecule that is already there — perhaps an antigen that appears on one type of cell and not on another¹. But in the Bertozzi approach, targeting also depends on factors (such as a deficiency or surfeit of a biosynthetic enzyme) that determine the extent to which the surface can be decorated with reactive sugars. This opens up new strategies for discriminating between cellular targets — strategies that depend not just on structure or composition but also on function and metabolic state. A cell can be targeted not only because of what it looks like, but also because of what it is doing. Whether one is interested in imaging or in therapy, the ability to accomplish function-based targeting is likely to be an important asset⁶.

Insights into how cellular physiology changes over time can also be extracted from these experiments. Because cell-surface

carbohydrates are made continuously, it is normally very difficult to distinguish newly synthesized variants from the large background of similar species already decorating the cell surface. But here, one labels only the newly synthesized variants, and if the chemical tag is designed properly one can use it to purify (or at least enrich) those variants for careful structural analysis. The challenges involved in such analyses are substantial but not insurmountable.

Finally, irrespective of its biological relevance, the method introduced by Bertozzi and colleagues is remarkable as a chemical process. As synthetic chemistry has advanced over the decades, the complexity of its targets and the specificity of its methods have developed in concert. From small molecules to large molecules, to molecular assemblies, materials and cells, chemists have tackled successively larger and more complex synthetic challenges⁷. The fact that specific chemical transformations can now be accomplished with spatial and temporal control in live animals is a major step forward for chemistry. ■

David A. Tirrell is in the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: tirrell@caltech.edu

- Guillemand, V. & Saragovi, H. U. *Curr. Cancer Drug Targets* **4**, 313–326 (2004).
- Lu, Y. J., Segal, E., Leamon, C. P. & Low, P. S. *Adv. Drug Delivery Rev.* **56**, 1161–1176 (2004).
- Prescher, J. A., Dube, D. H. & Bertozzi, C. R. *Nature* **430**, 873–877 (2004).
- Saxon, E. & Bertozzi, C. R. *Science* **287**, 2007–2010 (2000).
- Lowe, J. B. & Marth, J. D. *Annu. Rev. Biochem.* **72**, 643–691 (2003).
- Li, W.-H., Fraser, S. E. & Meade, T. J. *J. Am. Chem. Soc.* **121**, 1413–1414 (1999).
- Ruben, M. et al. *Angew. Chem. Int. Edn* **43**, 3644–3662 (2004).

A worthy adversary for malaria

Paul M. O'Neill

A remarkable set of antimalarial drug candidates has been developed by an international collaboration of scientists, using the age-old Chinese herbal medicine artemisinin as a template.

Nearly two billion people live in areas where malaria is endemic, and the incidence of this disease is increasing dramatically, mainly because many malaria-parasite strains have become resistant to the available drugs¹. The devastating spread of parasite resistance has prompted the worldwide search for new classes of effective antimalarial drugs. On page 900 of this issue, Vennerstrom and colleagues² describe the development of a new class of synthetic drug related to the natural antimalarial product artemisinin.

Artemisinin has been used in traditional Chinese herbal fever remedies for more than 1,500 years³. It is extracted from sweet wormwood (*Artemisia annua*), and is used as the basis for synthesizing several modern antimalarial drugs, such as artesunate and artemether. These artemisinin derivatives are increasingly important in the treatment of drug-resistant malaria, as they are the most potent antimalarials available, rapidly killing all blood stages of the malaria parasite *Plasmodium falciparum*. But the overall yield of the artemisinin extraction process is poor, so that its derivatives, though inexpensive by the standards of developed countries, are relatively expensive when put in the African context. In addition, the drugs only act for a short time, so the dosage regimen can lead to patient non-compliance and subsequent treatment failure. These shortcomings have prompted medicinal chemists to try to make entirely synthetic analogues of artemisinin with improved antimalarial properties⁴.

Artemisinins contain a structural feature known as an endoperoxide bridge (Fig. 1a) that is key to their antimalarial activity, and one of the major challenges in the pursuit of synthetic analogues has been to introduce this bridge into candidate drugs. Incorporating the endoperoxide 'warhead' is now possible, but many of the candidate analogues produced have significant drawbacks, including poor antimalarial activity and syntheses that are not stereoselective or not amenable to being scaled up^{5,6}. So, although many research groups had focused on the problem for some fifteen years, the goal of a cheap, synthetic, endoperoxide-based antimalarial had not been realized.

Vennerstrom *et al.*² now provide the solutions to many of these problems. They compiled a wish-list of ideal drug properties for the target molecule and stuck strictly to it through a process of drug-optimization to

create a new class of endoperoxide antimalarials that have superior properties to the artemisinin derivatives. The team consisted of chemists, parasitologists, pharmacokineticists and toxicologists.

The first stage of their drug discovery process involved a systematic examination of a chemical group known as secondary ozonides (or 1,2,4-trioxolanes; Fig. 1b). This class of compounds is well known to organic chemists as intermediates obtained by exposing alkenes to ozone. Although they have the requisite endoperoxide bridge, these molecules tend to be highly unstable—not a promising starting point, one might think. And unsurprisingly, some of the initial ozonide compounds tested turned out to be very poor antimalarials.

In artemisinin, the sensitive peroxide bridge is protected by bulky chemical rings, and the team attempted to mimic this feature by fusing a large, bulky group known as an adamantane ring onto the standard ozonide

ring system (Fig. 1b). This was their first breakthrough. Remarkably, not only were the resulting compounds stable, but when tested against human strains of the malaria parasite, they also had superior parasite-killing properties even to artesunate and artemether.

The first series of the compounds were poorly soluble in water, however, so the next step was to increase the compounds' water solubility so that they could be more easily absorbed from the gastrointestinal tract⁷. Many potential drugs containing chemical groups that increase water solubility were examined, and this led to the discovery of a candidate, OZ277 (Fig. 1c), that was well absorbed when administered orally to animals, and had outstanding antimalarial properties both *in vitro* and *in vivo*. Unlike the available artemisinin derivatives, OZ277 is structurally simple and its synthesis can be scaled up in a way that is economically feasible. What's more, its toxicological profiles are satisfactory. Following the development of a large-scale synthesis for OZ277, it has just entered 'first into man' studies.

Given the effectiveness of this new structurally simple drug, the question arises: how does it kill parasites? It seems probable that it shares a common mechanism of action with the artemisinins, where it is proposed that an interaction of the endoperoxide bridge with ferrous iron or haem from the red blood cell

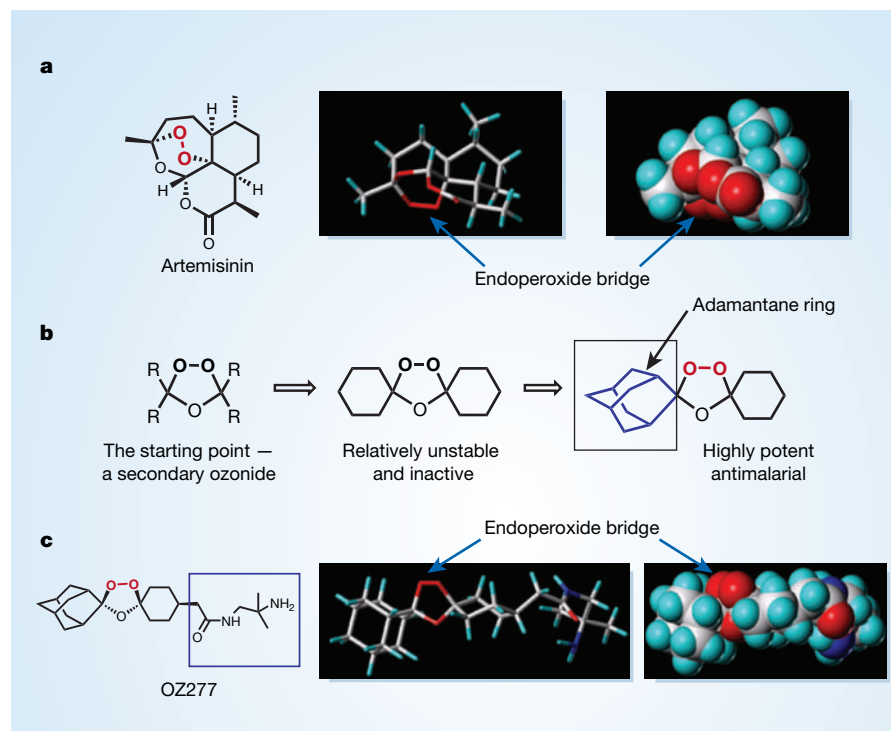


Figure 1 Artemisinin and its analogues. **a**, Chemical structure of artemisinin. The space-filled model on the right shows how the sensitive endoperoxide bridge is shielded by the ring system. **b**, Drug prototypes. Vennerstrom and colleagues² fused an adamantane ring system onto the basic ozonide ring to produce a new class of antimalarial drug with superior activity to artemisinin. **c**, The clinical candidate, OZ277. The adamantane system protects the endoperoxide bridge, and the boxed chemical group is key to the compound's improved water solubility and pharmaceutical properties. Carbon atoms, grey; hydrogen atoms, light blue; oxygen atoms, red; nitrogen atoms, dark blue.

in which the parasite lodges generates toxic, carbon-centred free radicals^{8,9}. Theoretically, these highly reactive molecules can modify key parasite proteins, disabling essential biological targets and killing the parasite. Using a technique known as spin-trapping, Vennerstrom and colleagues provide evidence that trioxolanes can indeed generate a carbon-centred free radical in a manner reminiscent of the artemisinins. There are several proposed protein targets for the noxious free radicals produced by artemisinin, one of which is an enzyme known as PfATP6 (ref. 10). Future studies should reveal whether OZ277 and its chemical siblings target the same proteins as the artemisinin derivatives.

The development of OZ277 is a flagship project for the Medicines for Malaria Venture^{11,12}. It is an excellent example of how a well-managed partnership between academia and major pharmaceutical companies can have a significant impact on antimalarial product research and development.

Basing the drug development process on a chemically unstable entity such as a secondary ozonide was a daring move. And the research that enabled ozonides to be redesigned, not only to increase the chemical and metabolic stability, but also to provide phenomenal antimalarial properties, is impressive. The subsequent tailoring of the 'ozonide' molecule to enhance its availability to the body was hugely successful — the new synthetic analogues are more potent and act for longer *in vivo* than artemether and artesunate by some margin. As such, when combined with a second antimalarial, this new class could offer the best solution to date for destroying drug-resistant malaria parasites. ■

Paul M. O'Neill is in the Departments of Chemistry and of Pharmacology, University of Liverpool, Liverpool L69 7ZD, UK.

e-mail: p.m.oneill01@liverpool.ac.uk

1. Winstanley, P. A., Ward, S. A. & Snow, R. W. *Microbes Infect.* **4**, 157–164 (2002).
2. Vennerstrom, J. L. *et al. Nature* **430**, 900–904 (2004).
3. Klayman, D. L. *Science* **228**, 1049–1055 (1985).
4. Borstnik, K., Paik, I. H., Shapiro, T. A. & Posner, G. H. *Int. J. Parasitol.* **32**, 1661–1667 (2002).
5. O'Neill, P. M. & Posner, G. H. *J. Medicin. Chem.* **47**, 2945–2964 (2004).
6. Vroman, J. A., Alvim-Gaston, M. & Avery, M. A. *Curr. Pharm. Design* **5**, 101–138 (1999).
7. Haynes, R. K. *Curr. Opin. Infect. Dis.* **14**, 719–726 (2001).
8. Posner, G. H. & Oh, C. H. *J. Am. Chem. Soc.* **114**, 8328–8329 (1992).
9. Wu, Y. K. *Acc. Chem. Res.* **35**, 255–259 (2002).
10. Eckstein-Ludwig, U. *et al. Nature* **424**, 957–961 (2003).
11. Vennerstrom, J. L., Dong, Y., Chollet, J. & Matile, H. US Patent 6,486,199 (2002).
12. www.mmv.org/pages/page_main.htm

Nature Outlook: Malaria

More about the challenges posed by malaria appears in the collection of articles published as a *Nature Outlook* supplement, beginning on page 923 of this issue. The articles cover scientific, social and political problems, with the emphasis on Africa.

Condensed-matter physics

Vortices weave a tangled web

David R. Nelson

In high-temperature superconductors, quantized vortex filaments can be twisted up into a DNA-like double helix. An experiment is proposed to test how easily these vortex lines cut through each other.

Inside a superconductor, electrical currents flow without resistance. Almost as remarkable as this electron flow without dissipation are the quantized, thread-like vortices of charge that swirl like miniature tornadoes around lines of magnetic field. Last year, Alexei Abrikosov shared the Nobel Prize in Physics for his brilliant 1957 prediction (made well before similar developments in high-energy theory and astrophysics) that, in a class of materials called 'type II' superconductors, a regular lattice of parallel vortex filaments, aligned with an external magnetic field, would form¹. In high-temperature copper-oxide superconductors, discovered 30 years later, Abrikosov's vortex lattice actually 'melts' over an appreciable range of magnetic field and temperature^{2,3}. Olson Reichhardt and Hastings⁴, writing in *Physical Review Letters*, now propose a key experiment that could unravel the physics of these vortices as they become entangled in the melted state.

A typical phase diagram for a high-temperature superconductor is shown in Fig. 1, overleaf, as a function of temperature and of the magnetic field induced in the material⁵. The Meissner phase of the diagram, in which surface currents completely exclude an applied magnetic field, follows the horizontal axis (where there is no magnetic induction in the superconductor) and terminates at the temperature at which the material ceases to be a superconductor — its transition temperature. According to Abrikosov's original theory, an applied magnetic field would penetrate the material as a regular lattice of quantized vortex filaments when the magnetic induction is greater than zero and below some critical field line (Fig. 1). But thermal fluctuations melt this lattice above the 'coexistence region' — where liquid and solid phases coexist at slightly different densities⁶.

What is this melted liquid of fluctuating, randomly braided vortex filaments like? An important property is its shear viscosity, which determines the ease with which point-like, linear or planar defects can pin the vortices in place. These defects might be missing atoms (for example, oxygen vacancies), columnar damage tracks from heavy-ion radiation, or more extended objects such as grain boundaries or globules of a different atomic phase. The pinning of vortex lines, thereby immobilizing them relative to the host material, is crucial in many high-field

applications. Indeed, if vortices are allowed to move in response to an applied current (which exerts a force on them), their motion dissipates energy, and resistanceless flow in a magnetic field is impossible. A vortex crystal resists shear deformations, so that pinning the Abrikosov flux lattice is relatively easy, like tacking a carpet to a slippery floor. However, pinning the melted lattice in a vortex liquid is much more difficult — this requires a high viscosity, similar to the property that allows us (temporarily, at least) to nail a pancake of 'silly putty' to a wall. Silly putty is composed of long, tangled polymer chains, and it is tempting to think of a melted liquid of Abrikosov vortices (Fig. 1) as a related 'directed polymer melt' of entangled vortex filaments. Crucial to the high viscosity of real polymer melts, however, are large energy barriers to the polymer chains crossing each other.

Do vortices, which are singularities in the underlying superconducting order, behave as impenetrable lines, or do they cut freely through one another? The answer depends on field and temperature. The energy barriers against flux cutting are very small near the critical field line (Fig. 1), where quantized vortices first form. However, when nearly parallel vortex lines cross, the 'quantum of vorticity' effectively doubles. The quantum of vorticity is a universal combination of Planck's constant, the speed of light and the electron charge, and determines the strength of the swirling supercurrents around a vortex. Well below the critical field line, this doubling leads to a very large crossing energy. Two entangled vortices can always cut each other easily if the lines deform so as to cross each other at a large angle, but this is resisted by the line tension. Although there is some experimental evidence for a large viscosity in a vortex liquid, theoretical estimates of the ratio of the melting temperature and the crossing energy at that temperature vary widely^{6,7}.

Olson Reichhardt and Hastings⁴ propose an elegant way to probe that all-important vortex crossing energy, using a magnetic force microscope (MFM). An MFM is similar to an atomic force microscope, except that it has a magnetized tip that pulls on the end of an individual vortex line⁸. It should be possible to pin the entry points of two vortex lines in place, and then use the MFM to wind one vortex around another into a double helix (Fig. 1). If the energy barriers to

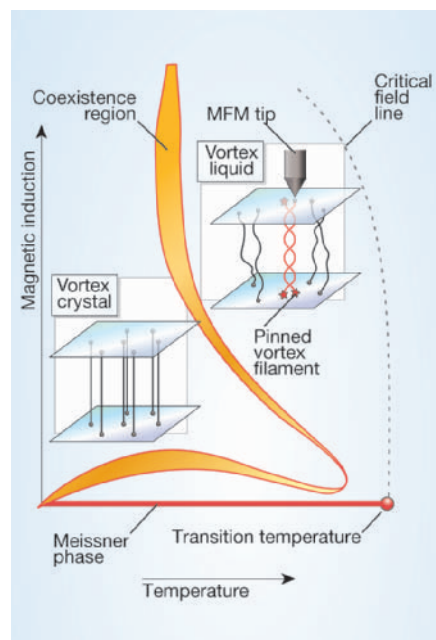


Figure 1 Equilibrium phase diagram for a high-temperature superconductor. The axes represent temperature and the induced magnetic field in the superconductor. For zero magnetic induction, the superconductor is in the Meissner phase up to its transition temperature. But when there is magnetic induction, vortex filaments appear in the material, represented here as lines around which supercurrents swirl. Superconducting order, interrupted by vortex lines, exists below the critical field line. At low temperature and below the coexistence region (in which both liquid and solid phases are possible), the vortex lines form an ordered Abrikosov vortex lattice; at higher temperatures above the coexistence region, the lattice melts to become an entangled vortex liquid. Olson Reichhardt and Hastings⁴ propose that the vortices in this phase could be manipulated, even wound up into double helices, using a magnetic force microscope (MFM).

flux-cutting are large, the vortex pair will twist up like a rubber band attached to the propeller of an old-fashioned model airplane. Both the applied force and the net displacement of the vortex tip can be monitored and compared with theoretical predictions⁴. In practice, flux-cutting will eventually occur, causing a highly twisted pair to relax and thus limit the build-up of stored energy. The mechanical braiding of a vortex pair would create additional supercurrents; these would occur in a pattern resembling a compressed version of the field lines generated by a solenoidal, or barrel-like, electromagnet. By monitoring the twist and subsequent relaxation with an MFM, the dynamical cutting of vortex lines could be probed in some detail.

MFM experiments on vortex lines might seem reminiscent of recent experiments in biophysics⁹. Among the striking observations are that 'supercoiled' DNA, created through helix crossings, can undergo relax-

ation mediated by a remarkable enzyme called a topoisomerase¹⁰. Using an MFM to detach a vortex line from a columnar pin in the superconductor¹¹ would be like tearing apart the DNA double helix¹². Just as biophysics experiments on single molecules provide an alternative to the averaging over large numbers of molecules that characterizes traditional biochemistry, MFM experiments on individual vortices are an attractive alternative to macroscopic probes of vortex physics.

For instance, the abrupt freezing transition of the vortex liquid that occurs on cooling does not occur in copper-oxide superconductors if the magnetic field is high¹³. One explanation for this puzzling behaviour is that point pinning, resulting from intrinsic disorder such as oxygen vacancies in the material, produces a new, disordered 'vortex glass' ground state at low temperatures¹⁴. Alternatively, if flux-cutting barriers are high, the dense, tangled vortices may simply drop out of equilibrium when cooled and form a 'polymer glass' of high viscosity⁶. Careful macroscopic measurements on samples of different thickness might be

able to distinguish between these scenarios. However, if realized, the MFM experiment of Olson Reichhardt and Hastings could get right to the heart of the matter, by probing individual flux lines directly.

David R. Nelson is in the Department of Physics, and in the Division of Engineering and Applied Science, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: nelson@cmt.harvard.edu

- Tomlin, S. *Nature* **425**, 548 (2003).
- Safar, H. *et al.* *Phys. Rev. Lett.* **69**, 3370–3373 (1992).
- Zeldov, E. *et al.* *Nature* **375**, 373–377 (1995).
- Olson Reichhardt, C. J. & Hastings, M. B. *Phys. Rev. Lett.* **92**, 157002 (2004).
- Crabtree, C. W. & Nelson, D. R. *Physics Today* April, 38–45 (1997).
- Nelson, D. R. *Defects and Geometry in Condensed Matter Physics* Chs 7, 8 (Cambridge Univ. Press, 2002).
- Carraro, C. & Fisher, D. S. *Phys. Rev. B* **51**, 534–538 (1995).
- Wadas, S., Fritz, O., Hug, H. J. & Guentherodt, H.-J. *Z. Phys. B* **88**, 317–320 (1992).
- Bustamante, C., Bryant, Z. & Smith, S. B. *Nature* **421**, 423–427 (2003).
- Strick, T., Croquette, V. & Bensimon, D. *Nature* **404**, 901–904 (2000).
- Hatano, N. & Nelson, D. R. *Phys. Rev. B* **56**, 8651–8673 (1997).
- Essevez-Roulet, B., Bockelmann, U. & Heslot, F. *Proc. Natl Acad. Sci. USA* **94**, 11935–11939 (1997).
- López, D. *et al.* *Phys. Rev. Lett.* **80**, 1070–1073 (1998).
- Fisher, D. S., Fisher, M. P. A. & Huse, D. *Phys. Rev. B* **43**, 130–159 (1991).

Cell division

Timing the machine

Bruce Bowerman

During cell division everything must happen at the right time, or errors occur. A common cellular control device, protein phosphorylation, is now shown to time the assembly of a key part of the division machinery.

Cells divide and thereby multiply. This fundamental process is central to the development and survival of all organisms, and mistakes in it are responsible for a plethora of human diseases, from Down's syndrome to cancer. Accordingly, cell division — also known as mitosis — has received much attention from biologists. This attention has led to the discovery and analysis of a cycle of events that influences key regulatory proteins¹, but the mechanisms by which these proteins in turn influence the machinery of mitosis are less well understood. Writing on page 908 of this issue, Mishima *et al.*² help to mitigate this disparity, describing direct links between cell-cycle regulators and the cell-division machinery.

The central machine in cell division is the bipolar mitotic spindle, an apparatus that partitions the duplicated genome of a mother cell equally into two daughter cells¹. The spindle is composed largely of microtubules — relatively rigid but highly dynamic tubes formed by the polymerization of tubulin proteins. Microtubules are nucleated by two microtubule-organizing centres (also called centrosomes in animal cells), one at each

spindle pole (Fig. 1). Spindle microtubules, in conjunction with associated proteins, capture and separate the duplicated genome, which comes in the form of long DNA molecules known as sister chromatids.

Spindle assembly begins during genome duplication, when the single centrosome that a cell inherits at birth also duplicates. The resulting two centrosomes migrate apart, grow, and nucleate more microtubules, which radiate out in all directions. The growing, or 'plus', ends of some microtubules capture pairs of sister chromatids, each sister at first remaining bound to its duplicate. Eventually, all sister chromatids are captured, with the two sisters in each pair connected to opposite poles. Subsequently, the protein-based glue between paired sisters is dissolved, and poleward forces move them apart — a stage of mitosis known as anaphase.

During anaphase a remarkable transition in spindle structure occurs. Many of the microtubules projecting from each pole do not capture chromatids, and some instead interdigitate, as their plus ends grow from opposite poles and pass each other. During anaphase, these 'antiparallel' microtubules

are gathered into bundles, in a region of overlap midway between the two poles, forming a structure that is referred to variously as the spindle midzone, the spindle interzone or the central spindle (Fig. 1). This intriguing structure seems to be important late in mitosis, promoting the ingression of a membrane furrow that ultimately partitions the dividing cell into separate daughters — a process called cytokinesis³.

Studies of both roundworm (*Caenorhabditis elegans*) and mammalian cells have shown that assembly of the central spindle requires a two-protein complex dubbed centralspindlin⁴. One of its protein constituents is a member of the kinesin family of motor proteins, called ZEN-4 in *C. elegans* and MKLP1 in humans. The other is a signalling protein of the Rho family — CYK-4 in *C. elegans*, MgcRacGap in mammals. These two proteins form a complex that crosslinks microtubules of opposite polarity. When functional centralspindlin is lost, the central spindle does not assemble and cytokinesis is defective.

Although the requirement for centralspindlin is well established, how its cross-linking activity is restricted to anaphase, and the functional significance of this restriction, has remained unknown. Mishima *et al.*² now present data suggesting that a cell-cycle regulator adds phosphate groups to (phosphorylates) centralspindlin to prevent assembly of the central spindle before anaphase. Dephosphorylation by another regulator then promotes assembly during anaphase.

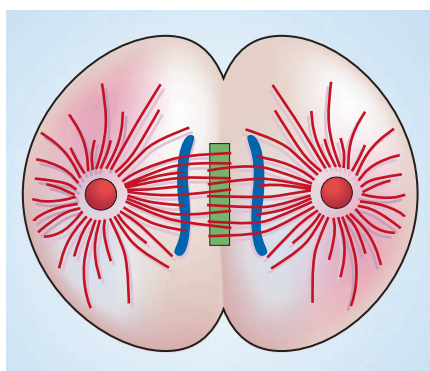


Figure 1 The mechanics of cell division. The mitotic spindle in a dividing animal cell is composed of microtubules (red lines) projecting from each of two spindle poles (also known as centrosomes; red circles). Some microtubules capture sister chromatids (blue), shown separating during the anaphase period of the cell cycle. Some microtubules interdigitate midway between the two spindle poles, where they are cross-linked through the activity of the centralspindlin complex to form the central spindle (green box). Mishima *et al.*² have shown how the phosphorylation and dephosphorylation of a key component of centralspindlin ensures that the central spindle assembles at the correct time.

Kinesins, such as ZEN-4 and MKLP1, are typically dimers that bind microtubules through two head regions, which can 'walk' along a single microtubule⁵. The head regions might also allow centralspindlin complexes to bundle microtubules, by binding to and crosslinking antiparallel microtubules. A kinesin's head regions are joined to neck regions, and Mishima *et al.* focus on amino acids that are found in the necks of both ZEN-4 and MKLP1 and that look like targets for phosphorylation.

The presence of these conserved potential phosphorylation sites, and the ability of the neck region to promote motor activity *in vitro*, compelled Mishima *et al.* to test the functional importance of neck phosphorylation. Their data show that a cell-cycle regulator called Cdk1–cyclin B phosphorylates the neck regions of both proteins. To examine the consequences of this phosphorylation, the authors performed both *in vitro* and *in vivo* tests. Kinesins use chemical energy obtained from the hydrolysis of adenosine triphosphate (ATP) molecules to power their movement on microtubules. Phosphorylation of the ZEN-4/MKLP1 neck greatly lowered this ATP hydrolysis activity *in vitro*, and reduced the affinity of the proteins for microtubules. So, as long as Cdk1–cyclin B is active, microtubule crosslinking is prevented. This would delay the assembly of the central spindle until anaphase, when Cdk1–cyclin B is known to be inactivated.

To address the functional significance of this phosphorylation *in vivo*, Mishima *et al.* expressed an altered MKLP1 in cultured mammalian cells. In this altered protein the amino acid alanine, which cannot be phosphorylated, replaces amino acids that would normally be targeted by Cdk1–cyclin B. This alteration resulted in premature bundling of microtubules before anaphase, and interfered with sister-chromatid segregation. So, neck phosphorylation does indeed seem to prevent premature assembly of the central spindle, and might thereby facilitate the proper capture and segregation of sister chromatids.

To complete their story, the authors also found that a cell-cycle phosphatase called CDC14 can dephosphorylate the ZEN-4/MKLP1 necks. In mutant *C. elegans* cells lacking functional CDC14, ZEN-4 failed to localize to the central spindle during anaphase, presumably because the protein's neck could not be dephosphorylated. In contrast, as would be expected if neck phosphorylation prevents central-spindle assembly, ZEN-4 that had been altered by alanine substitution to prevent neck phosphorylation did localize to the central spindle when CDC14 was not functional. Thus, by targeting the neck region of the centralspindlin kinesin, cell-cycle kinase and phosphatase activities can restrict assembly of the central spindle to anaphase.



100 YEARS AGO

Inaugural address by the Right Hon. A. J. Balfour: Reflections suggested by the New Theory of Matter.

If we jump over the century which separates 1804 from 1904, and attempt to give in outline the world-picture as it now presents itself to some leaders of contemporary speculation, we shall find that in the interval it has been modified, not merely by such far-reaching discoveries as the atomic and molecular composition of ordinary matter, the kinetic theory of gases, and the laws of the conservation and dissipation of energy, but by the more and more important part which electricity and the ether occupy in any representation of ultimate physical reality... But to-day there are those who regard gross matter, the matter of everyday experience, as the mere appearance of which electricity is the physical basis; who think that the elementary atom of the chemist, itself far beyond the limits of direct perception, is but a connected system of monads or sub-atoms which are not electrified matter, but are electricity itself... Surely we have here a very extraordinary revolution. Two centuries ago electricity seemed but a scientific toy. It is now thought by many to constitute the reality of which matter is but the sensible expression.

From *Nature* 18 August 1904.

50 YEARS AGO

Institution of Electronics Exhibition in Manchester. The commercial section was impressive chiefly for its evidence of steady progress in the development of known techniques. Applications of television were prominent, and there were on view no fewer than three industrial closed-circuit television channels... The Institute of Cancer Research exhibited ultrasonic echo-locating equipment, used for the examination of brain structure. When the equipment is in use, a quartz transducer in acoustic contact with the head emits a narrow, pulsed beam of ultrasound. Any echoes incident on the transducer are transmitted as pulses through amplifier and display circuits, and appear on an oscillograph screen. The position of the pulse on the screen gives a measure of the distance of the source of the echo from the transducer. At present, observations are being compared with what is known of the 'normal' brain, with the view of the possible identification of abnormal structures.

From *Nature* 21 August 1954.

Mishima and colleagues' extensive analysis² documents a direct link between cell-cycle regulatory proteins and a key change in the structure of the mitotic spindle. Nevertheless, data just published by another group⁶ indicate that our understanding remains incomplete. Surprisingly, these investigators show that deleting the CDC14 gene in *C. elegans* does not result in detectable defects in assembly of the central spindle, chromosome segregation or cytokinesis. Instead, mutant worms lacking CDC14 undergo extra, non-lethal cell divisions during larval development, apparently because a protein that is unrelated to ZEN-4 cannot be dephosphorylated. At least in *C. elegans*, then, CDC14 is not required for

cell division. One must conclude that dephosphorylation is either not necessary for cell division or can be accomplished without CDC14. No doubt we can expect further chapters in this intriguing story. ■

Bruce Bowerman is at the Institute of Molecular Biology, 1370 Franklin Boulevard, University of Oregon, Eugene, Oregon 97403, USA.

e-mail: bbowerman@molbio.uoregon.edu

1. Scholey, J. M., Brust-Mascher, I. & Mogilner, A. *Nature* **422**, 746–752 (2003).
2. Mishima, M., Pavicic, V., Grüneberg, U., Nigg, E. A. & Glotzer, M. *Nature* **430**, 908–913 (2004).
3. Glotzer, M. *Annu. Rev. Cell Dev. Biol.* **17**, 297–300 (2001).
4. Mishima, M., Kaitna, S. & Glotzer, M. *Dev. Cell* **2**, 41–54 (2002).
5. Schliwa, M. & Woehlke, G. *Nature* **422**, 759–765 (2003).
6. Saito, R. M., Perreault, A., Peach, B., Satterlee, J. S. & van den Heuvel, S. *Nature Cell Biol.* doi:10.1038/ncb1154 (2004).

Palaeoclimatology

Fresh angle on the polar seesaw

Trond M. Dokken and Kerim H. Nisancioglu

During the last glacial period, climatic variation in the Northern and Southern Hemispheres was evidently linked. Modelling work points to freshwater discharge into the North Atlantic as a driving factor.

Climate during the period from 60,000 to 25,000 years ago, referred to as Marine Isotope Stage 3, was exceptionally variable. Ice-core records from Greenland suggest that the Northern Hemisphere underwent a series of rapid warming episodes, each followed by gradual cooling. The Southern Hemisphere had a similar climate pattern, but with relatively slow warming and less extreme temperatures. These so-called Dansgaard–Oeschger (DO) events occurred several times during this period, but the physical processes behind the timing and amplitude of the recorded temperature changes are unclear. On page 851 of this issue, Knutti and colleagues¹ present a conceptual model suggesting that discharges of fresh water into the North Atlantic had a larger part to play than was previously thought.

The relative timing of the climate changes in each hemisphere can be investigated by using variations in methane content to synchronize the timescales of changes in the ice cores taken from Greenland and Antarctica. And it seems that the temperature swings in the two hemispheres were offset, with changes in Antarctica preceding those in Greenland by just over a millennium². This out-of-phase relationship between the two hemispheres has been invoked as evidence for a bipolar climate 'seesaw'³, where cooling in the Northern Hemisphere is balanced by warming in the Southern Hemisphere (Fig. 1).

In general, the ocean's meridional overturning circulation (MOC) transports heat northward in the dense salty water near the surface of the Atlantic. As the water approaches the Norwegian–Greenland seas

it cools and sinks into the deep ocean, where it is carried back south. According to the classical seesaw model, DO events are caused by a disruption of this circulation, resulting in cooling in high northern latitudes and a build-up of heat in the Southern Hemisphere^{4,5}. One prime suspect for the cause of this disruption is a freshening of the surface ocean in the North Atlantic and the Nordic seas caused by massive releases of melt water from the surrounding ice sheets. Such a freshening would reduce the density of the water and prevent the northward limb of the MOC from sinking. The MOC would recover once the surface density increased again, and the Northern Hemisphere would then warm rapidly, drawing heat from the Southern Hemisphere.

The classical seesaw model assumes an anti-phase relationship between temperature changes in the north and south, with no time lag. But studies of Antarctic and Greenland ice cores⁶ and model simulations⁷ suggest that the anti-phase changes in Antarctic temperature lag those recorded in Greenland by 400–800 years. Based on these results, Stocker and Johnsen⁸ introduced a revised model — the thermal bipolar seesaw — in which temperatures in the two hemispheres are out of phase, but with a lead of Greenland over Antarctica. In this model, the Southern Ocean acts as a heat reservoir that takes time to warm or cool, with a timescale that matches the estimated lag between the hemispheres.

But simulations based on the thermal seesaw model, although improved, still fail to match the temperature record deduced from the ice cores. So Knutti and colleagues¹ have

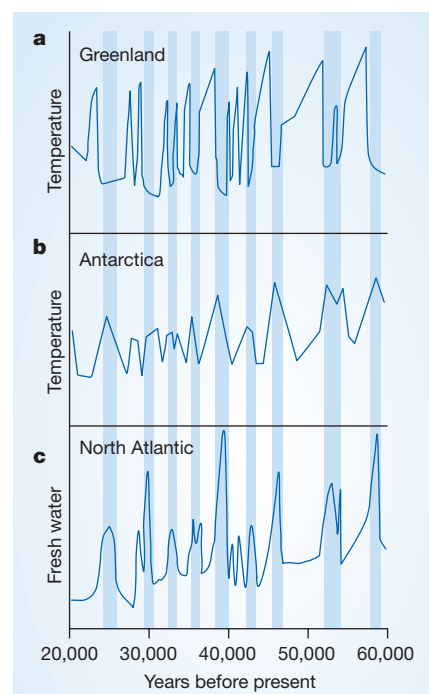


Figure 1 Abrupt climate changes in the past glacial period. a,b, Simplified graphs of the temperature data retrieved from the Greenland (GRIP; a) and Antarctic (Byrd; b) ice cores. Increases in temperature in the Northern Hemisphere are balanced by decreases in the Southern Hemisphere, although the swings are slightly out of synch. c, A smoothed graph showing discharges of fresh water into the North Atlantic from the surrounding ice sheets, inferred from ocean sediment cores that contain glacial deposits¹⁵. There is a clear correspondence between periods of increased melt water in the North Atlantic and the most pronounced warming episodes in the Southern Hemisphere. This provides support for a northern trigger for the climate changes — as is assumed by the various bipolar climate seesaw models.

now updated this model to improve the timing and shape of the simulated temperatures. They used a climate model that simulates processes in both the ocean and the atmosphere, and find that, in addition to disturbing the MOC, fresh water from Northern Hemisphere ice sheets might have amplified the temperature response in the Southern Hemisphere by another, more direct mechanism. They find evidence of anomalous transport of heat from the North to the South Atlantic Ocean in response to a large freshwater perturbation at high northern latitudes.

To account for this, the authors propose a mechanism that invokes a fast 'wave-adjustment', by which the depth of the ocean's upper layer of warm water (thermocline water) fluctuates by a process somewhat resembling a tide. This adjustment in the Atlantic sets up a southward cross-equatorial ocean current, which transports heat from north to south during periods of massive ice-sheet dis-

charge. The authors suggest that up to a third of the southern temperature signal is due to this current, and that the remaining two-thirds is associated with the thermal-seesaw effect and changes in the large-scale MOC.

The classical bipolar seesaw and Knutti and colleagues' revised thermal-freshwater seesaw are intriguing, as they present physically based models to explain a set of observations. And indeed, the modelled temperature results fit very well with the ice-core temperature changes (Fig. 5 on page 855). To test all aspects of the new concept, however, more evidence is necessary, for example data on the strength of the overturning circulation. An obvious test for future studies would be to see how much of the variability observed in Antarctic temperature data can be explained by freshwater data from the North Atlantic.

The various versions of the bipolar seesaw model assume that climate changes in the Northern Hemisphere trigger a response in the Southern Hemisphere. And some evidence for a northern trigger is provided by the fact that observed increases in fresh water discharged into the North Atlantic follow the pattern predicted by the models relative to the ice-core temperature data (Fig. 1). However, an increasing number of calculations suggest that Antarctic temperature changes precede those in Greenland by 1,000–2,000 years^{2,9}. Therefore, an alternative theory is that the trigger lies in the Southern Hemisphere. Model experiments^{10–12} and ocean sediment-core data^{13,14} suggest that a variety of processes in the Southern Hemisphere might have provoked changes in the MOC. These include changes in the strength of westerly winds and the circumpolar current; changes in Southern Ocean density structure; and gradual warming triggered by a shift in the main source of water entering the South Atlantic, either via the warm Indian Ocean or the cold Pacific Ocean.

For now, the notion of a southern trigger for climate changes is an interesting theory that lacks a conceptual model able to explain all the observations from Antarctica and Greenland. Regardless of whether north or south leads in DO events, we need to understand better why shifts in the MOC occur. However, conflicting evidence and numerous diverse lines of argument on how the climate of the two hemispheres is linked confuse the issue. At present, we lack the necessary data from the northern and southern oceans to put palaeoceanographic constraints on the past history of MOC mode switches from this bipolar perspective. In any event, understanding the cause and effect of previous abrupt climate changes is crucial for a rational assessment of the probability of such events occurring in the future. ■

Trond M. Dokken and Kerim H. Nisancioglu are at the Bjerknes Centre for Climate Research, University of Bergen,

Allégaten 55, Bergen 5007, Norway.
e-mail: trond.dokken@bjerknes.uib.no

1. Knutti, R., Flüchiger, J., Stocker, T. F. & Timmermann, A. *Nature* **430**, 851–856 (2004).
2. Blunier, T. et al. *Nature* **394**, 739–743 (1998).
3. Broecker, W. S. *Paleoceanography* **13**, 119–121 (1998).
4. Crowley, T. J. *Paleoceanography* **7**, 489–497 (1992).
5. Stocker, T. F. *Science* **282**, 61–62 (1998).
6. Steig, E. J. & Alley, R. B. *Ann. Glaciol.* **35**, 451–456 (2002).
7. Schmittner, A. et al. *Quat. Sci. Rev.* **22**, 659–671 (2003).

8. Stocker, T. F. & Johnsen, S. J. *Paleoceanography* **18**, doi:10.1029/2003PA000920 (2003).
9. Hinno, L. A. et al. *Quat. Sci. Rev.* **21**, 1213–1228 (2002).
10. Toggweiler, J. R. & Samuels, B. *Deep-Sea Res.* **142**, 477–500 (1995).
11. Knorr, G. & Lohmann, G. *Nature* **424**, 532–536 (2003).
12. Weaver, A. J. et al. *Science* **299**, 1709–1713 (2003).
13. Ninnemann, U. S. & Charles, C. D. *Earth Planet. Sci. Lett.* **201**, 383–396 (2002).
14. Peeters, F. J. C. et al. *Nature* **430**, 661–665 (2004).
15. Dokken, T. M. & Jansen, E. *Nature* **401**, 458–461 (1999).

Structural biology

Anthrax hijacks host receptor

James G. Bann and Scott J. Hultgren

An atomic picture of how anthrax toxin binds to its host's cells reveals that the toxin commandeers a host receptor protein and tricks it into helping the toxin enter the cell.

In 2001, *Bacillus anthracis* made headlines when US Senators Thomas Daschle and Patrick Leahy received letters containing anthrax spores, highlighting the urgent need for an effective treatment against the bacterium. Once exposed to *B. anthracis*, the only treatment available involves a 60-day course of antibiotics that have unpleasant side-effects¹. The race to develop more palatable alternatives that will work at any stage of infection is now focusing on anthrax toxin, the protein complex responsible for the bacterium's lethal effects.

On page 905 of this issue, Liddington and colleagues² report the X-ray crystal structure of one of the anthrax toxin proteins, the protective antigen (PA), bound to its receptor from the host's cell, capillary morphogenesis

protein 2 (CMG2). This work explains the structural basis of how anthrax toxin recognizes CMG2, and suggests a mechanism by which CMG2 is duped into behaving as a molecular switch that controls the transfer of anthrax toxin into the cell's cytosol, an event that ultimately proves fatal to the host.

Anthrax toxin is composed of three proteins: protective antigen (so named because it is used as a vaccine), oedema factor and lethal factor. PA is a large protein consisting of four domains (I–IV), primarily involved in targeting the toxin to host cells by recognizing CMG2. The crystal structure² reveals that the high-affinity binding of PA with CMG2 (ref. 3) is due partly to the involvement of a magnesium ion at the interface between them. A key aspartic acid residue

Oceanography

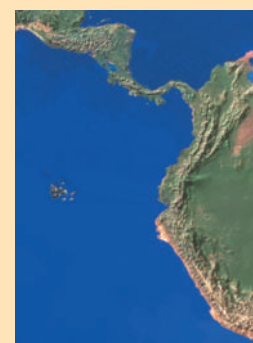
Islands in the stream

During the *Beagle's* visit to the Galapagos Islands in 1835, Charles Darwin noted that the local climate was far less warm than would be expected from the islands' position on the Equator. The air-conditioning effect is due to the cooling influence of the surrounding oceans — part of which, according to C. Eden and A. Timmermann (*Geophys. Res. Lett.* **31**, L15308; 2004), arises from the very presence of the islands.

As this satellite image shows, the Galapagos are isolated in the vastness of the Pacific Ocean, lying about 1,000 km west of South America. This is nonetheless an oceanographically sensitive location, because the islands

obstruct two components of a system of wind-driven ocean currents in the equatorial Pacific. The cool Southern Equatorial Current flows westwards as part of the Pacific subtropical gyre, and splits into a northern and a southern branch at the Galapagos. The subsurface Equatorial Undercurrent transports water eastwards between and beneath these two branches, and almost stops dead where it hits the islands.

Using a high-resolution numerical model, Eden and Timmermann have simulated equatorial Pacific currents with and without the Galapagos topography. The differences are significant. The islands produce



a wake-like pattern in both currents, with flow anomalies extending up to 2,000 km in an east–west direction. And as a result of stronger upwelling of cooler water from depth, sea surface temperatures just west of the Galapagos are up to 2 °C lower than they would otherwise be — hence the comparatively temperate climate. Heike Langenberg

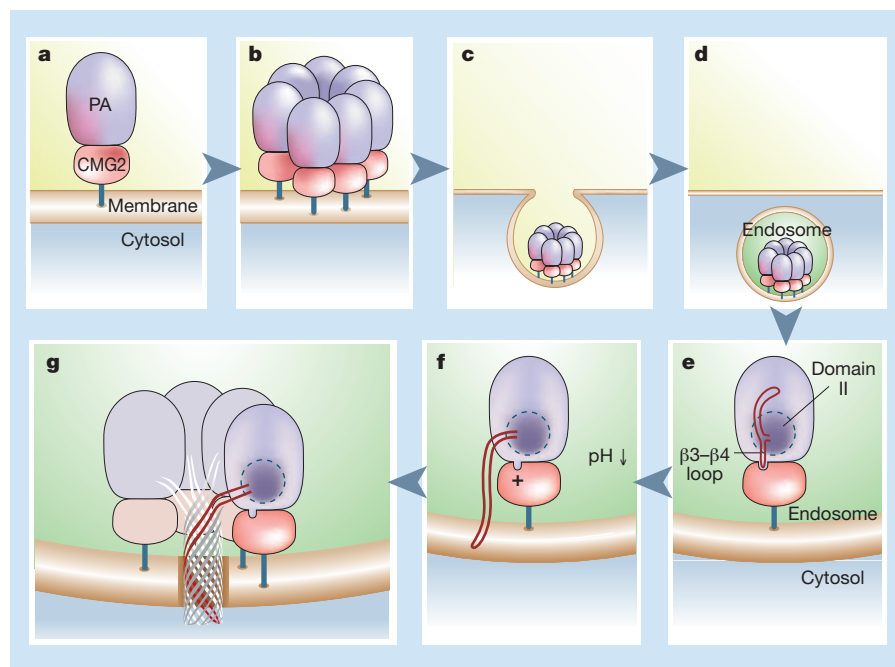


Figure 1 Throwing the anthrax switch. Liddington and colleagues² suggest how anthrax toxin tricks a host cell receptor into helping it into the cell. **a**, The anthrax protective antigen (PA) binds the host capillary morphogenesis protein 2 (CMG2) on the outside of the cell. **b**, Seven PA-CMG2 modules link together to form a heptameric complex, which binds to the other anthrax toxin proteins, oedema factor and lethal factor (not shown). **c**, A deep pocket forms in the cell membrane. **d**, The neck of the pocket is pinched off, leaving the toxin-receptor complex inside the endosome. **e**, **f**, The molecular switch. For simplicity, a single PA-CMG2 monomer is shown. When PA binds to CMG2 (**e**), a loop of PA domain II ($\beta 3$ - $\beta 4$) is gripped in a groove on the CMG2 surface. Inside the endosome (**f**) the pH decreases, generating a positive charge in the CMG2 groove. This repels the $\beta 3$ - $\beta 4$ loop, resulting in a large conformational change in PA domain II. The loop and some neighbouring strands peel away and insert into the endosome membrane. **g**, They twist around strands from neighbouring PA-CMG2 modules to form a pore, allowing the oedema factor and the lethal factor into the host cytosol.

in domain IV of PA works in conjunction with a metal-ion-dependent adhesion site (MIDAS) on CMG2 to coordinate the ion.

An atomic structure of CMG2 (ref. 4) revealed that it is very similar to a domain in proteins called integrins, which mediate the attachments between cells and the extracellular matrix. So a fascinating feature of the new crystal structure² is the discovery that PA does indeed recognize CMG2 using a similar mechanism to the one by which extracellular matrix proteins bind to integrins. Specifically, PA binds to CMG2 in a manner similar to the way in which extracellular matrix protein type IV collagen recognizes $\alpha 2 \beta 1$ integrin⁵: the collagen uses an aspartic acid to help coordinate a magnesium ion together with a MIDAS site on the integrin. But this raises a question: if integrins and CMG2 are so similar structurally, how does the anthrax toxin tell them apart? Unexpectedly, the crystal structure² shows that domain II of PA has a small β -hairpin loop ($\beta 3$ - $\beta 4$) that fits snugly into a groove on the CMG2 surface. Integrins do not have a comparable groove, explaining how PA is able to discriminate between them and CMG2.

Once PA binds to CMG2 on the host-cell surface, a protease clips PA in two. The smaller portion diffuses away, and the larger part

remains bound to the CMG2 receptor, eventually forming a complex of seven PA-CMG2 modules, called a pre-pore⁶. The oedema factor and/or the lethal factor bind to this PA-CMG2 complex, triggering a process called endocytosis, by which the PA-CMG2 complex is engulfed into the cell (Fig. 1). The area of the cell membrane containing the toxin-receptor complex forms a deep pocket into the cell. The neck of the pocket is pinched off to create a bubble-like organelle, an endosome, with the toxin-receptor complex inside, still attached to the membrane. To inject the oedema factor and the lethal factor into cells, the seven PA molecules must act together to form a straw-like structure — a pore — bridging the endosome membrane and opening out into the cell cytosol (Fig. 5 on page 907). The pore transfers the oedema factor and the lethal factor to the cytosol, leading ultimately to cell death through the disruption of vital physiological processes^{7,8}.

Liddington and colleagues' crystal structure² reveals a molecular-switching mechanism in the complex that might control the formation of this pore (Fig. 1). The groove on CMG2 that interacts with PA domain II contains a crucial residue (histidine 121) that holds the PA in the right conformation until

it is ready to insert into the endosome membrane. But what throws this molecular switch so that the toxin can enter the cell? The authors propose that the answer might be in the pH of the local environment. Their model (Fig. 1e-g) suggests that once the endosome is formed, the internal pH decreases and histidine 121 is protonated, becoming positively charged. This repels a nearby arginine on PA, reducing the affinity of the $\beta 3$ - $\beta 4$ loop of PA for CMG2. Consequently, the PA domain II undergoes a large conformational change, with the $\beta 2$ - $\beta 3$ strands adjacent to the $\beta 3$ - $\beta 4$ loop peeling away from PA like the skin of a banana peeling away from the fruit. The $\beta 2$ - $\beta 3$ strands are lined with several histidines, and protonation of these probably helps this unwrapping process⁹. Once free of CMG2 and PA, the strands insert into the endosome membrane and form the pore by twisting around the strands from the six neighbouring PA molecules⁹. Essentially, CMG2 acts as a pH-sensitive switch, holding the PA in the right shape until just the right time, before releasing it to form the pore.

CMG2 was discovered only recently, and it is proposed to have a role in the assembly of the basement membrane, the meshwork of extracellular matrix proteins that helps to support cells¹⁰. Although the normal biological function of CMG2 is as yet unclear, presumably it is not to facilitate translocation of anthrax toxin into the cell. Rather, Liddington and colleagues' analysis² suggests that anthrax toxin hijacks CMG2, employing it as a molecular switch to help release the toxin into the cell. This structure will provide a good starting point for evaluating the energetics and mechanism of pore formation, enabling the design of drugs aimed at derailing the critical early steps of anthrax function. It could also provide clues to how other pore-forming toxins, such as α -haemolysin from *Staphylococcus aureus*, undergo such large conformational changes¹¹.

James G. Bann is in the Department of Chemistry, Wichita State University, Wichita, Kansas 67260-0051, USA.

Scott J. Hultgren is in the Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.

e-mails: jim.bann@wichita.edu
hultgren@borcim.wustl.edu

1. Jeffers, M. D. et al. *Emerg. Infect. Dis.* **8**, 1138-1144 (2002).
2. Santelli, E., Bankston, L. A., Leppla, S. H. & Liddington, R. C. *Nature* **430**, 905-908 (2004).
3. Wigelsworth, D. J. et al. *J. Biol. Chem.* **279**, 23349-23356 (2004).
4. Lacy, D. B. et al. *Proc. Natl Acad. Sci. USA* **101**, 6367-6372 (2004).
5. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. *Cell* **101**, 47-56 (2000).
6. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. *Nature* **385**, 833-838 (1997).
7. Leppla, S. H. *Proc. Natl Acad. Sci. USA* **79**, 3162-3166 (1982).
8. Park, J. M., Greden, F. R., Li, Z. W. & Karin, M. *Science* **297**, 2048-2051 (2002).
9. Blaustein, R. O., Koehler, T. M., Collier, R. J. & Finkelstein, A. *Proc. Natl Acad. Sci. USA* **86**, 2209-2213 (1989).
10. Bell, S. E. et al. *J. Cell Sci.* **114**, 2755-2773 (2001).
11. Galdiero, S. & Gouaux, E. *Protein Sci.* **13**, 1503-1511 (2004).

Obituary

Francis Crick (1916–2004)

Alexander Rich and Charles F. Stevens, respectively an early collaborator of Crick's and a long-standing colleague at the Salk Institute, describe the life and work of one of the great thinkers of twentieth-century biology.

Alexander Rich writes:

Francis Crick, who died on 28 July at the age of 88, was trained as a physicist but became arguably the most influential biologist of the twentieth century. His great curiosity was coupled to highly original thinking; through force of intellect he obtained answers to many fundamental problems. In seminars he often demanded clarity from speakers, thereby generating some tension. However, he had a lively sense of humour, sharp but never malicious. Crick had no PhD students and only a rare postdoctoral fellow, but nonetheless often worked closely with a collaborator. Above all, he was a very kind and considerate person.

Born in Northampton and trained at University College London, Crick started graduate work in physics at the beginning of the Second World War. His "unimaginably dull" thesis project was to define the viscosity of water at high pressures. In a career-altering episode during the Battle of Britain in 1940, a bomb fell through the roof of the physics laboratory and exploded on his instrument. Crick then went to work for the Admiralty, designing 'clever' mines. He started biological research in 1947, working initially in the Strangeways Laboratory in Cambridge, where he devised experiments to measure the viscosity of cytoplasm. This left him somewhat dissatisfied, and in 1949 he joined Max Perutz at Cambridge's Cavendish Laboratory, investigating protein structure for his PhD.

Biological research in the late 1940s was moving in several different directions, but making little progress. A central, unsolved problem was how genetic information is transmitted from an organism to its offspring. There was little awareness in the community at large that this problem could be attacked at the molecular level, and most scientists thought that genes were proteins. In the mid-1940s, Oswald Avery and colleagues had presented evidence that DNA might be the hereditary material, but that conclusion was not widely accepted. What was needed was a catalytic event.

That event was the arrival of Jim Watson at Cambridge in 1951. Crick was then 35 years old and Watson 23, but both shared a passion for understanding the molecular



basis of genetics. They were convinced that DNA was the genetic material.

What happened next is widely known. Crick's familiarity with the X-ray diffraction patterns produced by helical structures, the access to DNA diffraction patterns taken by Rosalind Franklin, and Watson's intuitive attempts to pair nucleotide bases, facilitated by Jerry Donohue's critical intervention regarding their correct structure, led to the double-helical model of DNA in an astonishing few weeks. Their method, largely adopted from Linus Pauling, involved using accurate metal skeletal atoms to assemble

a double helix, with its component chains running in opposite directions and joined by complementary base pairs in the centre. The complementarity of the two strands in the structure provided a mechanism for inheritance, in that each single strand could act as a template for assembling its complement — leading to two identical duplex molecules. The information is in the sequence of the bases. The significance of this work was not widely recognized at first, but after a few years the steady accumulation of new evidence for the double helix made it apparent that this was a



Watson and Crick in 1953.

transforming milestone in the development of biological science.

The next question was: how could the information housed in DNA be used to produce proteins, which do the work of the cell? Many thought that RNA played a role, but nothing was known of its structure, let alone its function. How could a nucleic acid such as RNA determine the specific sequence of amino acids that makes up a protein? This was called the coding problem, but it is interesting that many biologists at the time were largely unaware that there even was a problem.

Naïve early attempts were made by Crick, Watson, Leslie Orgel and myself to formulate RNA foldings that might have specific amino-acid-binding pockets. The first published proposal for a solution, however, was by George Gamow, a colourful and highly talented theoretical physicist. Gamow playfully took the lead in forming the 'RNA Tie Club', with 20 members, one for each naturally occurring amino acid, and four honorary members, one for each nucleotide base. In addition to having a striking tie, members sent each other monographs about how the code could be solved.

In a paper circulated in the mid-1950s, Crick pointed out that nucleic acids seem to associate naturally with other nucleic acids. Thus he proposed that there might be 20 classes of 'adaptor' RNA molecules, which could line up along a template nucleic acid and each bind to a specific amino acid. Although most people were sceptical, such molecules, now called transfer RNAs, were soon discovered by Mahlon Hoagland and Paul Zamecnik. Thus, by logical deduction and intuition,

Crick uncovered a key link between the RNA copy of DNA (messenger RNA) and the amino acids in protein synthesis.

In 1955, Crick invited me to the Cavendish to work on RNA fibres, and to stay with him and his artist wife Odile at their house in Portugal Place. They enjoyed hosting parties in their third-floor sitting room; the atmosphere at such gatherings was lively, with many jokes and good humour. Doing science in the mid-1950s was fun, with few worries about funding, and the exciting prospect of new discoveries on all sides.

Indeed, my short visit there extended to more than six months, because a newly arrived issue of *Nature* reported a novel form of an amino acid polymer, polyglycine II. We decided to try to solve its structure using molecular models. After only four hours of work, the coiled structure we built predicted the intensity and spacing of the published X-ray diagram. Crick suggested that we might try to write this up quickly and see if we could get it published in the next week's *Nature*. But then he paused, because he thought that the authors might feel badly, so we invited them over to look at the structure. It was characteristic of Crick that he was sensitive to people's feelings and would not intentionally cause them distress.

We later recognized that if we took three hydrogen-bonded strands from the polyglycine lattice, we could twist them slightly to make a coiled-coil structure, which was a model for collagen — the long fibrous protein of skin. Optical diffraction studies demonstrated that it was the correct structure.

This close collaboration made me appreciate the force of Crick's intellectual drive and the subtlety of his thinking. Our research progressed through an endless dialogue, looking at many sides of the problem. Crick had a strong competitive approach to science — other groups were working on collagen. But his basic attitude was not ego-driven; it had deeper roots. Like Pauling, my postdoctoral mentor, Crick was motivated to show that living systems could be explained by chemistry and physics, thereby supporting his world view as an atheist.

There remained the problem of determining the number of nucleotide bases that are needed to specify one amino acid for protein synthesis. This problem was solved in 1961 by Crick and Sydney Brenner, a collaborator of Crick's for many years. In a microbial experiment, a mutagen was used that added nucleotides. Adding one or two nucleotides blocked protein synthesis, but after three nucleotide bases were added, protein synthesis resumed. This simple but elegant experiment showed that the genetic code involves triplets of bases.

Certain inconsistencies arose in interpreting the interactions between the triplet of bases that defines an amino acid in messenger RNA and the triplet of bases in the transfer RNA molecule. To account for that, in 1966 Crick proposed the 'wobble' hypothesis, in which one base of a transfer RNA could adopt two different positions, hydrogen bonding in two different ways. This led to the complete genetic code, relating each amino acid to one or more nucleotide triplets. These monumental discoveries provided the basic framework for understanding the flow of information, and defined the major features of all living systems at the molecular level.

In 1977, Francis moved to the Salk Institute in La Jolla, California, home of his long-time collaborator Orgel. The move was associated with a shift in his interests away from problems of molecular biology and towards brain mechanisms — more specifically, consciousness. I stayed in touch with Francis continuously, and in our last conversation, about a week before his death, he said he was feeling much the same and working hard on a manuscript that he hoped to publish. His concern at the time was that the paper was too long, and that the journal might want it to be considerably reduced.

Alexander Rich is in the Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA.

Charles F. Stevens continues:

A week before Crick died, Terry Sejnowski and I went to see him at home to talk about plans for the new Crick-Jacobs Center for Computational and Theoretical Biology at the Salk Institute. We found him, surrounded by papers, sitting in a chair next to a window that looked out over his patio. He looked much the same as usual, dressed in slacks, a shirt open at the neck, a sports jacket. But a walking stick leaned against his chair, and his ankles were swollen.

We talked for about an hour, most of the time about his new passion, the claustrum. Crick was writing a review article on this obscure brain nucleus, and he had sent me a rough draft the week before to get my comments. What had fascinated Crick was that the claustrum gets its input from many cortical areas in the brain, and sends its output back to those areas.

This arrangement made him think that perhaps the claustrum was a sort of conductor of the cortical 'orchestra'. He had a strong hunch, based on its connection pattern, that the claustrum might be a neural structure central to consciousness — we argued for a while about the likelihood of this possibility — and he said he hoped his article would stimulate research on a neglected brain area that he cared about. As we were leaving, Crick faltered briefly as he got up from his chair, and then said, with a characteristic twinkle in his eye as we shook hands, "I can still manage to stand up to say goodbye". He had had colon cancer for several years. His chemotherapy was no longer working, and he had said, quite dispassionately, that it was unlikely he would live through September. As ever, he was passionate about his science and unsentimental about what he could not control.

Many things in this last meeting were characteristic of Crick. Just as the secret of heredity lay in a structure, so did he seek gold once more — consciousness, this time — in brain structure. He always



Crick with Christof Koch in March this year.

sorted through problems to find those that could be formulated as crisp questions, just as now he defined what properties he thought brain structures involved in consciousness should have, and then browsed through possibilities to find answers. And he was really interested in qualia — how subjective feelings arise — but settled for a question that he thought could be answered, what neural structures and activities are required for consciousness. He had the idea that, in getting any answer, this might, if you're clever enough and lucky enough, give you insights that will help with the harder question you really want to answer.

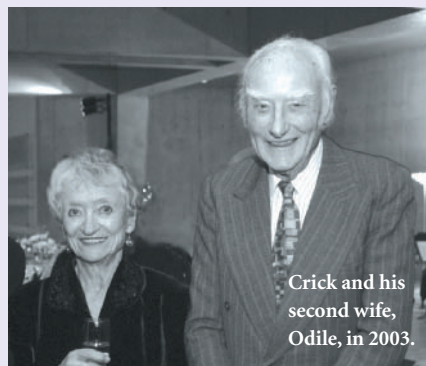
Crick was a theorist rather than an experimentalist, and he believed strongly that theory is necessary in biology not only to organize and explain phenomena, but also to define the questions that need to be answered. After defining such questions, he then stimulated (sometimes nagged) experimentalists to answer them. Although he was devoted to theory, generally his theoretical notions were not especially quantitative. Rather, he sought to abstract the essential and very simple mechanisms from the detail.

After he moved from Cambridge to the Salk a quarter of a century ago, Crick used to invite neurobiologists to spend time with him. I made the pilgrimage about 20 years ago, before I moved there, and we spent a week talking about the visual system and the hippocampus — all day, every day, and sometimes into the evening.

But Crick found certain people especially congenial for his give-and-take, and formed long-lasting and close collaborations that were particularly important to him. In neuroscience, first Graeme Mitchison and then, for the past 15 years, Christof Koch were his main collaborators. Koch regularly worked with him on ideas about consciousness, and Crick depended on these interchanges for formulating his programme to identify what they always called the NCC (the neural correlates of consciousness). And the give-and-take with Koch was terribly important emotionally as well as intellectually.

Francis stimulated many neurobiologists, myself included, by his keen questions about their work and his sharp insights. But his contributions to neurobiology vanish in comparison with what he did in molecular biology. Possibly his most influential contribution to neurobiology was making the study of consciousness respectable. Francis said, famously, about his work with Watson that, "It's true that by blundering about, we stumbled on gold, but the fact remains that we were looking for gold". Perhaps, had he been 20 or 30 years younger when he started in neurobiology, he might have found gold in the study of consciousness, too.

Charles F. Stevens is in the Molecular Neurobiology Laboratory, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: stevens@salk.edu



Crick and his second wife, Odile, in 2003.

Microbiology

How to stomach a stomach infection

Science **305**, 1003–1006 (2004)

Researchers might have discovered why most people infected with the stomach bug *Helicobacter pylori* never become ill. Masatomo Kawakubo *et al.* have found a naturally occurring antibiotic that might offer protection against the bacterium.

H. pylori infects around half of the world's population. It exists in the stomach lining and can cause ulcers, gastritis and malignant lymphoma. But most infected individuals stay symptom-free, suggesting that the body possesses some form of natural defence. The bacteria are rarely found in the deep layers of the stomach lining, where cells secrete mucus loaded with proteins called O-glycan glycoproteins, containing unique α 1,4-N-acetylglucosamine residues. So Kawakubo *et al.* decided to test these molecules for microbicidal activity.

The authors found that O-glycans suppress the growth and mobility of *H. pylori* by interfering with its ability to form a key cell-wall component. They also work on a variety of *H. pylori* strains. Because the antibiotic is naturally produced, the authors hope that their discovery will boost the development of safer drugs to prevent and treat *H. pylori* infections.

Helen Plicher

Chemistry

Azaspiracid all joined up

Angew. Chem. Int. Edn **43**, doi:10.1002/anie.200460695 & 200460696 (2004)

Azaspiracid-1 is a complex neurotoxin, originally isolated from the mussel *Mytilus edulis* (pictured), that can cause acute poisoning in humans. But its mode of action is poorly understood — not least because of uncertainties about its chemical structure, which has been one of the hottest challenges in synthetic chemistry. The structure is fiendishly complex, with nine interlinked rings, made of carbon, oxygen and nitrogen atoms, arranged in highly specific orientations that have been difficult to reproduce in the laboratory.

Now, however, K. C. Nicolaou *et al.* report a total chemical synthesis of azaspiracid-1. By comparing the synthetic molecule with the natural product, they are able to confirm the structure and so rule out an earlier proposal of a slightly different arrangement. The key step in the synthesis, say the authors, came when a flaw in the original proposal was tracked down to a portion of the molecule with a similar structure to lissoketal, a much simpler marine natural product whose



structure was already well understood.

The ability to prepare azaspiracid-1 in sufficient quantities allows further investigation of what makes the molecule so toxic. The authors also hope that it will open the route for the development of a test to detect it in seafood and water.

Mark Peplow

Neurobiology

Night visionaries

Curr. Biol. **14**, 1309–1318 (2004)

Working at night can be advantageous for a bee — many plants flower nocturnally, and there are fewer predators around. But bees are famously visual creatures, navigating with the aid of familiar landmarks. So how do nocturnal species such as the South American sweat bee *Megalopta genalis* find their way around in the dark?

The answer, report Eric J. Warrant and colleagues, is that they have acquired a suite of sophisticated modifications to the standard equipment. Like daytime bees, *M. genalis* has 'apposition' eyes made up of separate lenses that each focus onto an area of light-sensitive cells called a rhabdom. But the total rhabdom area is much larger than that of daytime bees, helping to provide 30 times more light-detecting power.

Even this is not enough for the bees to find their way home after dark, however — they typically live in 6-millimetre-wide holes in the rainforest undergrowth. It turns out that the bees' optical nerves are also interconnected to facilitate 'spatial summation', combining the signals from the few photons they detect to build a cumulative low-resolution picture. The authors argue that the bees probably use temporal summation too. The resulting picture is probably both blurry and jerky — but when you're in the forest at night, it's better than nothing.

Michael Hopkin

Energy generation

Power from heat and noise

Appl. Phys. Lett. **85**, 1085–1087 (2004)

Onboard power for spacecraft must be lightweight and efficient. Currently, it is typically provided either by fuel cells or by thermoelectric devices, which convert a temperature gradient (produced by a

radioactive source) directly into electrical power. But S. Backhaus *et al.* think they can do better.

They have made a thermoacoustic generator that converts heat to electricity with around 18% efficiency, more than twice as good as thermoelectrics. The device uses the same principle as a Stirling heat engine: expansion of a hot gas creates the engine's power stroke. In the present case, motion of the piston is driven by expansion of helium inside a tube that is hot at one end and cool at the other. This motion is converted to electricity using linear alternators — coils of copper wire attached to the piston that move through a magnetic field.

The pressure oscillation of the gas basically corresponds to an acoustic travelling wave inside the engine, which is amplified by the temperature gradient (provided here by electrical heating, although it could come from radioactive decay for space applications). This amplification allows the motion to arise purely out of thermal noise, just like acoustic feedback in a public-address system.

Phillip Ball

Genetics

Advantage, testis

Proc. Natl Acad. Sci. USA **101**, 11695–11700 (2004)

In males with the 'juvenile spermatogonial depletion' mutation, sperm are produced but don't develop, rendering adults infertile. Jan Rohozinski and Colin E. Bishop offer insight into the basis of this disorder, as well as a proposal for the evolution of male germline function.

In male mice with this mutation, the authors identified an error in the *Utp14b* gene. This gene encodes a protein required for the production of 18S ribosomal RNA, which in turn is needed for general protein synthesis. The authors found that *Utp14b* is expressed primarily in male germline cells, explaining why the mutation is so detrimental to sperm production.

The authors also identified a related gene, *Utp14a*, on the X chromosome; this gene is expressed in non-germline cells and in a later stage of sperm production. Curiously, the sequence of *Utp14b* indicates that it was derived from spliced *Utp14a* messenger RNA that jumped back into the genome.

So why are there two copies of *Utp14*? Rohozinski and Bishop propose that *Utp14b* acquired testis-specific expression, and that this property proved evolutionarily advantageous, as the highly proliferative germ line would require high levels of protein synthesis. In support of this view, humans have a similar 'retrotransposon' that seems to have arisen independently of the mouse version — suggesting that it, too, was selected by evolutionary pressure.

Angela K. Eggleston

Quantum teleportation across the Danube

A real-world experiment marks a step towards worldwide quantum communication.

Efficient long-distance quantum teleportation¹ is crucial for quantum communication and quantum networking schemes². Here we describe the high-fidelity teleportation of photons over a distance of 600 metres across the River Danube in Vienna, with the optimal efficiency that can be achieved using linear optics. Our result is a step towards the implementation of a quantum repeater³, which will enable pure entanglement to be shared between distant parties in a public environment and eventually on a world-wide scale.

Quantum teleportation is based on a quantum channel, here established through a pair of polarization-entangled photons shared between Alice and Bob (Fig. 1). We have implemented this by using an 800-metre-long optical fibre installed in a public sewer system located in a tunnel underneath the River Danube, where it is exposed to temperature fluctuations and other environmental factors.

For Alice to be able to transfer the unknown polarization state of an input photon $|\chi\rangle_b$, she has to perform a joint Bell-state measurement on the input photon b and her member, c , of the shared entangled photon pair (c and d). Our scheme allows her to identify two of the four Bell states, the optimum achievable with only linear optics^{4,5}.

As a result of this Bell-state measurement, Bob's 'receiver' photon d will be projected into a well defined state that already contains full information on the original input photon b , except for a rotation that depends on the specific Bell state that Alice observed. Our teleportation scheme therefore also includes active feed-forward of Alice's measurement results, which is achieved by means of a classical microwave channel together with a fast electro-optical modulator (EOM). It enables Bob to perform the unitary transformation on photon d to obtain an exact replica of Alice's input photon b .

Specifically, if Alice observes the $|\psi^-\rangle_{bc}$ Bell state, which is the same as the initial entangled state of photons c and d , then Bob already possesses the original input state. But if Alice observes the $|\psi^+\rangle_{bc}$ state, he introduces a π -phase shift between the horizontal and vertical polarization components of photon d by applying a voltage pulse of 3.7 kV on the EOM. For successful operation, Bob has to set the EOM correctly before photon d arrives. Because of the reduced velocity of light within the fibre-based quantum channel (two-thirds of that *in vacuo*), the classical signal arrives

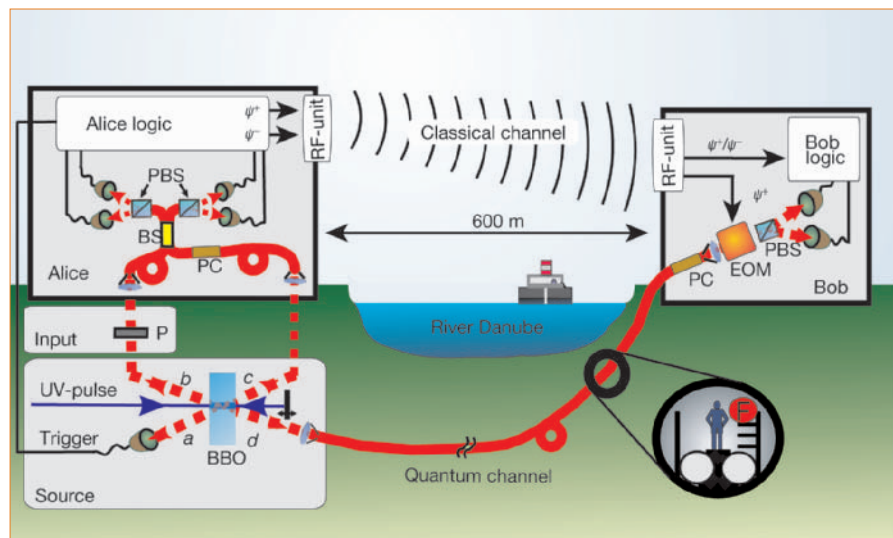


Figure 1 Long-distance quantum teleportation across the River Danube. The quantum channel (fibre F) rests in a sewage-pipe tunnel below the river in Vienna, while the classical microwave channel passes above it. A pulsed laser (wavelength, 394 nm; rate, 76 MHz) is used to pump a β -barium borate (BBO) crystal that generates the entangled photon pair c and d and photons a and b (wavelength, 788 nm) by spontaneous parametric down-conversion. The state of photon b after passage through polarizer P is the teleportation input; a serves as the trigger. Photons b and c are guided into a single-mode optical-fibre beam splitter (BS) connected to polarizing beam splitters (PBS) for Bell-state measurement. Polarization rotation in the fibres is corrected by polarization controllers (PC) before each run of measurements. The logic electronics identify the Bell state as either $|\psi^-\rangle_{bc}$ or $|\psi^+\rangle_{bc}$ and convey the result through the microwave channel (RF unit) to Bob's electro-optical modulator (EOM) to transform photon d into the input state of photon b .

about 1.5 microseconds before the photon.

We demonstrated the teleportation of three distinct polarization states: linear at 45°, left-handed circular or horizontal. The teleportation fidelity achieved was 0.84, 0.86 or 0.90 for the 45°, for each of these input states, respectively. These fidelities comfortably surpass the classical limit of 0.66 (ref. 6) and prove that our teleportation system is operating correctly. Without operation of the EOM, however, Bob observes a completely mixed polarization for the 45° and circular polarization input states, causing the observed fidelity for these states to drop to 0.54 and 0.59, respectively, in the absence of active unitary transformation. The deviation from the random fidelity of 0.5 is due to statistical fluctuations in the observed counts.

Each measurement run lasted for 28 h and the rate of successful teleportation events was 0.04 per second. Polarization stability proved to be better than 10° on the fibre between Alice's and Bob's labs, corresponding to an ideal teleportation fidelity of 0.97 over a full measurement run. Hence, despite the exposure of our system to the environment, high-fidelity teleportation was still achievable without permanent readjustments.

We have demonstrated quantum teleportation over a long distance and with high fidelity under real-world conditions

outside a laboratory. Our system combines for the first time, to our knowledge, an improved Bell-state analyser with active unitary transformation, enabling a doubling of the efficiency of teleportation compared with earlier experiments based on independent photons^{7,8}. Our experiment demonstrates feed-forward of measurement results, which will be essential for linear-optics quantum computing^{9–11}, and constitutes a step towards the full-scale implementation of a quantum repeater.

Rupert Ursin*, Thomas Jennewein*†, Markus Aspelmeyer*, Rainer Kaltenbaek*, Michael Lindenthal*, Philip Walther*, Anton Zeilinger*†

*Institute for Experimental Physics, University of Vienna, 1090 Vienna, Austria
e-mail: rupert.ursin@univie.ac.at

†Austrian Academy of Science, 1090 Vienna, Austria

1. Bennett, C. H. *et al.* *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
2. Bose, S., Vedral, V. & Knight, P. L. *Phys. Rev. A* **57**, 822–829 (1998).
3. Briegel, H. J., Dür, W., Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
4. Lütkenhaus, N., Calsamiglia, J. & Suominen, K.-A. *Phys. Rev. A* **59**, 3295–3300 (1999).
5. Kim, Y.-H. *et al.* *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
6. Popescu, S. *Phys. Rev. Lett.* **72**, 797–799 (1994).
7. Bouwmeester, D. *et al.* *Nature* **390**, 575–579 (1997).
8. Marcikic, I., de Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. *Nature* **421**, 509–513 (2003).
9. Riebe, M. *et al.* *Nature* **429**, 734–737 (2004).
10. Barrett, M. D. *et al.* *Nature* **429**, 737–739 (2004).
11. Gottesmann, D. & Chuang, I. L. *Nature* **402**, 390–393 (1999).

Competing financial interests: declared none.

Biomechanics

Hydrodynamic function of the shark's tail

The tail of most sharks has an elongated upper lobe that differs from the externally symmetrical tail structure common among bony fishes, but the hydrodynamic purpose of this asymmetric tail shape is unclear^{1–3}. Here we quantify water flow patterns in the wakes of freely swimming dogfish sharks and find that they have a ring-within-a-ring vortex structure, in contrast to the single rings shed by symmetrical fish tails. The branched-ring vortex is generated by the inclined angle of the tail's trailing edge and by its motion at an angle to the horizontal body axis; the vortex directs water backwards and downwards, which may increase the shark's vertical manoeuvrability.

We used digital particle-image velocimetry^{4,5} to analyse the hydrodynamic function of the tail in four spiny dogfish (*Squalus acanthias*; mean body length, 69 cm) swimming in a flow tank equipped with a vertical laser-light sheet. By using high-speed video (250 frames per second) to obtain particle images, we were able to analyse the time-dependent features of wake flow from the tail: six pairs of images were recorded throughout a tailbeat for five tailbeats from each of four individuals.

Three centres of vorticity were detected (Fig. 1a). The smaller dorsal vortex ring

includes centres (1) and (2) and is contained within a larger vortex ring, which includes centres (1) and (3). Counter-rotating centres in the dorsal vortex ring produce jet A; a second jet B develops later as a result of flow induced by the ventral tip vortex. Jets A and B are directed on average at -35° to the freestream flow, before combining to produce a single broad jet C. Initially, little downstream flow is seen between same-sign vorticity centres (2) and (3) (Fig. 1a).

Velocity transects through the centres of vorticity confirm the presence of a small dorsal vortex ring linked to a larger vortex ring that has a diameter equivalent to the tail height (Fig. 1b,c). Circulation calculations show that mean vortex-centre (1) circulation ($0.0160 \text{ m}^2 \text{ s}^{-1}$) is not significantly different from the sum of negative circulation in vortex centres (2) and (3) (-0.0019 and -0.0165 , which is $-0.0184 \text{ m}^2 \text{ s}^{-1}$ in total; paired *t*-test, $P=0.3$), as required by Kelvin's law. These findings contrast with the single vortex rings generated by the symmetrical tail of the bony fishes^{6,7}. A strong posteroventrally directed jet flow is visible as a result of shark-tail motion (Fig. 1a,c), corroborating one classical model of shark-tail function^{2,8}.

We propose that the mechanism for generating the ring-within-a-ring vortex structure in swimming dogfish is similar

to that for fluid ejected from a pipe with an inclined opening (refs 9,10). Dorsal and ventral tip vortices (1) and (3) are shed as the tail beats from side to side. The dorsal lobe leads the ventral, so the tail presents an inclined edge to oncoming flow⁸. In a pipe with an inclined opening, a ring-within-a-ring vortex structure is generated owing to the temporal asymmetry in vortex roll-up between the longer and shorter edges^{9,10}.

Previous work on the symmetrical tail of the bony fishes^{6,7,11} has shown that the vortex wake consists of single-ring vortices with dimensions similar to tail height and tailbeat width. The difference in wake structure between sharks and bony fishes seems to be due to their different tail shape and motion. The effects of these wake morphology differences on manoeuvring performance and locomotor energetics remain to be determined.

C. D. Wilga*, G. V. Lauder†

*Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island 02881, USA
e-mail: cwilga@uri.edu

†Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

1. Grove, A. J. & Newell, G. E. *Ann. Mag. Nat. Hist.* **17**, 280–290 (1936).
2. Alexander, R. M. T. *J. Exp. Biol.* **43**, 131–138 (1965).
3. Thomson, K. S. *Paleobiology* **2**, 19 (1976).
4. Drucker, E. G. & Lauder, G. V. *J. Exp. Biol.* **202**, 2393–2412 (1999).
5. Wilga, C. D. & Lauder, G. V. *J. Exp. Biol.* **205**, 2365–2374 (2002).
6. Lauder, G. V. *Am. Zool.* **40**, 101–122 (2000).
7. Nauen, J. C. & Lauder, G. V. *J. Exp. Biol.* **205**, 1709–1724 (2002).
8. Ferry, L. A. & Lauder, G. V. *J. Exp. Biol.* **199**, 2253–2268 (1996).
9. Webster, D. R. & Longmire, E. K. *Phys. Fluids* **9**, 655–666 (1997).
10. Webster, D. R. & Longmire, E. K. *Phys. Fluids* **10**, 400–416 (1998).
11. Lauder, G. V. *Am. Zool.* **29**, 85–102 (1989).

Competing financial interests: declared none.

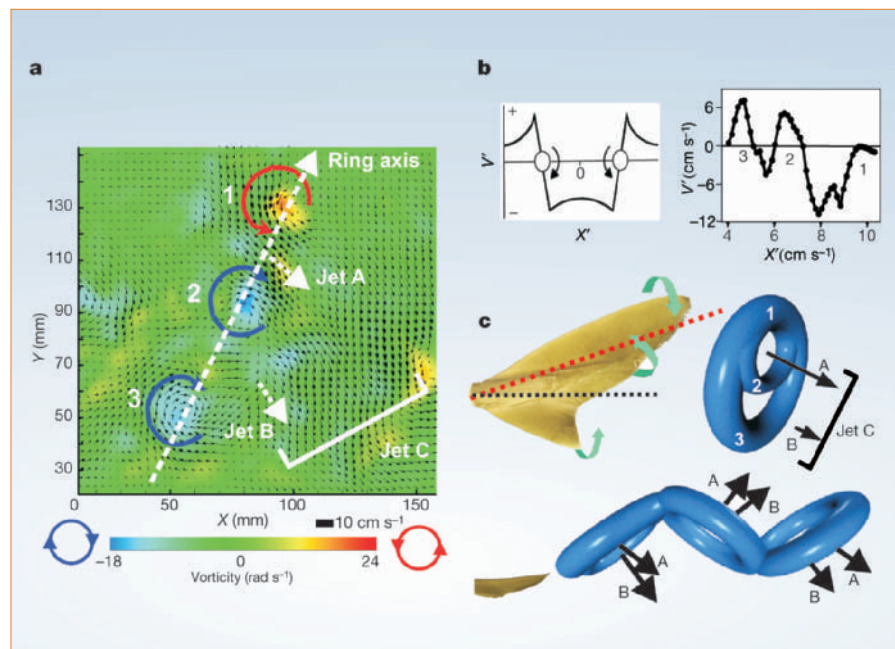


Figure 1 Digital particle-image velocimetry analysis, velocity profiles and vortex structures of rings shed from the tail of steadily swimming dogfish. **a**, Vorticity plot of a vertical slice through tail vortex rings; curved arrows indicate the three centres of vorticity. **b**, Theoretically predicted velocity distributions across a planar section of a vortex ring (left) and a similar plot based on empirical data from the sequence shown in **a** along the ring axis (right). X' , horizontal velocity; V' , vertical velocity (both relative to the ring axis). **c**, Side and top views of a shark vortex wake. The axis of tail rotation (red dotted line) is inclined to the horizontal axis of locomotion (black dotted line), generating a ring-within-a-ring vortex structure by a mechanism similar to a piston with an inclined exit orifice¹⁰. Successive tailbeats generate linked rings (bottom). Green arrows indicate dorsal and ventral tip vortices and roll-up of the vortex sheet along the trailing edge.

Strong hemispheric coupling of glacial climate through freshwater discharge and ocean circulation

R. Knutti¹, J. Flückiger¹, T. F. Stocker¹ & A. Timmermann²

¹Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

²IFM-GEOMAR, Leibniz-Institut für Meereswissenschaften, Düsternbrooker Weg 20, D-24105 Kiel, Germany

The climate of the last glacial period was extremely variable, characterized by abrupt warming events in the Northern Hemisphere, accompanied by slower temperature changes in Antarctica and variations of global sea level. It is generally accepted that this millennial-scale climate variability was caused by abrupt changes in the ocean thermohaline circulation. Here we use a coupled ocean–atmosphere–sea ice model to show that freshwater discharge into the North Atlantic Ocean, in addition to a reduction of the thermohaline circulation, has a direct effect on Southern Ocean temperature. The related anomalous oceanic southward heat transport arises from a zonal density gradient in the subtropical North Atlantic caused by a fast wave-adjustment process. We present an extended and quantitative bipolar seesaw concept that explains the timing and amplitude of Greenland and Antarctic temperature changes, the slow changes in Antarctic temperature and its similarity to sea level, as well as a possible time lag of sea level with respect to Antarctic temperature during Marine Isotope Stage 3.

The climate over much of the last glacial period was extremely variable on a millennial timescale. The North Atlantic climate was punctuated by warm phases recorded in proxies over most of the Northern Hemisphere¹. These so-called Dansgaard–Oeschger (DO) events² were characterized by changes from cold (stadial) towards warmer and wetter (interstadial) conditions, with shifts of up to 16 °C observed within a few decades in Greenland^{3–5}, followed by a more gradual cooling over a few hundred to thousand years. The long-lasting DO events were preceded by massive ice surges from the Northern Hemisphere ice sheets. These so-called Heinrich events are documented as thick layers of ice-rafted debris in marine sediments in the North Atlantic^{6,7}. They coincide with cold conditions in the North Atlantic region, warm episodes in Antarctica⁸ and with increases in sea level of 10 to 35 m (refs 9–12). Although the body of observational data is growing, the physical processes behind these events remain elusive.

Guided by evidence from marine sediments for massive reorganizations of the ocean circulation¹³, and modelling results^{14,15}, a concept known as the ‘thermal bipolar seesaw’ was suggested, in which abrupt changes in the strength of the ocean thermohaline circulation (THC) affect the polar climate through changes in the meridional heat transport^{14,16–20}. The THC is sensitive to the Atlantic freshwater balance and exhibits a threshold and hysteresis behaviour^{21–23}. It was argued that the increased glacial climate variability is a result of different stability properties of the glacial thermohaline circulation as compared to today¹⁹. Although in qualitative agreement with many proxies, important issues remain unclear from the published modelling studies. In particular, most models^{19,24} so far have underestimated either the large temperature shifts of 8 °C and more over Greenland and parts of the North Atlantic region, or the changes of about 3 °C over the Antarctic continent^{25–27}. Furthermore, it is difficult to reconcile the conspicuous temporal relationship between the abrupt shifts in Greenland² and the relatively slow changes in the Antarctic temperature reconstructions²⁸ in a physically consistent way.

Here we present results using the ECBILT-CLIO model^{29,30}, which consists of an ocean general circulation model, coupled to a sea-ice and a simplified dynamical atmosphere model. This coupled global atmosphere–ocean model (CGAOM) is forced by meltwater

discharge into the North Atlantic (see Methods). On the basis of the simulated climate events, we propose a mechanism by which freshwater discharge into the Northern Hemisphere directly affects the Southern Ocean temperature through two distinct, but intimately related, ocean circulation feedbacks. This concept, combined with a multi-millennial transient climate model simulation of abrupt climate events, provides a consistent framework to answer a number of open questions that have remained unresolved until now. This includes the timing, amplitude and spatial extent of DO and Heinrich events, the slow timescale associated with Antarctic warm phases, the similarity and lag of sea level to southern temperature and the source of the meltwater discharge.

The climate effect of freshwater discharge

To investigate the processes relating northern and southern high-latitude temperatures, three model simulations using idealized freshwater discharge scenarios are shown in Fig. 1. For case A, the freshwater discharge into the North Atlantic leads to a partial shutdown of the deepwater formation and a cooling in the North Atlantic region. The subsequent recovery to warm conditions coincides approximately with the cooling in the Southern Ocean region. For both cases B and C, the freshwater discharge is sufficient to stop the meridional overturning completely, causing a cooling in the north and a warming in the south. While the cooling in the North Atlantic has a similar magnitude in cases B and C, the Southern Ocean warming is larger for a larger freshwater input. This is in contradiction with the classical ‘thermal bipolar seesaw’ picture, in which southern temperature is determined by northern temperature only²⁰. In addition, southern temperature in case C starts to decrease several centuries before the onset of northern temperature increase. The similarity of southern temperature time series with the freshwater forcing in the North Atlantic is also simulated for different scenarios (not shown) and thus suggests a direct process by which northern freshwater discharge affects southern temperature.

To elucidate the physics of this process, equilibrium simulations using a sustained freshwater input into the North Atlantic of zero, 0.5 and 1.0 Sv (control, F0.5 and F1.0, respectively; 1 Sv = 10⁶ m³ s^{−1}) were performed. Figure 2a shows the simulated

temperature anomaly pattern in the Atlantic arising from the complete collapse of the meridional overturning simulated for F0.5. The surface North Atlantic and the deep Atlantic cool by several degrees, while the South Atlantic and Southern Ocean warm in the upper 1,500 m, consistent with previous model studies^{31–33}. A subsurface warming is seen north of about 60° N, caused by the cessation of deep convective mixing of cold surface water, in particular north of Iceland. Because fresh water affects both the THC and southern temperature as long as the THC is active, the effect of fresh water is best isolated when comparing the two states F0.5 and F1.0 in which the THC is collapsed. The additional freshwater discharge in F1.0 leads to an anomalous southward mass transport of about 2 Sv in the Atlantic south of 40° N, compensated by a return flow between 500 and 1,500 m in depth (Fig. 2b). This causes an anomalous southward heat transport of 0.1 PW, which warms the Southern Ocean sea surface temperature by about 1.5 °C. Although the subsurface Atlantic cools between 20° S and 40° N and warms north of 40° N, this is barely seen at the surface, probably owing to the strong stratification arising from the freshwater cap.

The anomalous meridional transport is due to a zonal density gradient caused by the freshwater input in the following way. The density changes associated with the North Atlantic fresh water trigger Kelvin waves, which propagate along the western Atlantic coast towards the Equator³⁴. Owing to the Coriolis effect, they are forced to travel along the Equator towards the coast of Africa, where they split into a northern and southern branch. While moving poleward, they radiate Rossby waves, which readjust the interior transport of the North and South Atlantic. Overall, our simulated global thermocline anomaly pattern (not shown) is consistent with this wave-adjustment mechanism^{34,35}. Rossby waves establish large-scale pressure (sea level) gradients that are accompanied by geostrophic flow anomalies (see Fig. 2c). North of the Equator these southward geostrophic currents are relatively strong, owing to the weakness of the Coriolis force and the presence of large sea-level gradients in the western boundary region. Cross-equatorial mass

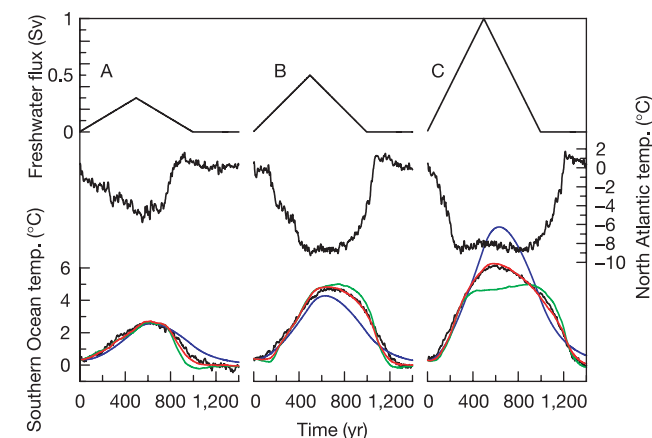


Figure 1 Temperature response to three scenarios of freshwater discharge into the North Atlantic. Freshwater discharge (top), 10-yr running mean of near-surface air temperature of the North Atlantic region (60° W–20° E, 50° N–80° N) (middle) and the Southern Ocean region (65° S–50° S) (bottom) are shown in black. The fresh water causes a partial shutdown of the THC in case A, and a complete shutdown in cases B and C. The classical thermal bipolar seesaw²⁰ (green) predicts the same amplitude of the southern warming for cases B and C and fails to explain the early cooling in southern temperatures. The extended thermal–freshwater seesaw concept (red) correctly predicts both an additional southern warming for case C that is related to the large freshwater input, and the slow southern temperature changes due to slow freshwater changes. The fit here explains more than 98% (r^2) of the southern temperature response simulated by the CGAOM. Fresh water alone (blue) is insufficient to explain the model response.

transport is possible, because the velocity anomalies occur close to the western boundary of the basin, where friction is important. Finally, a southward cross-equatorial heat transport of about 0.1 PW is established, in response to the North Atlantic freshwater flux anomalies. Typical response times are of the order of one or two decades³⁶. These conclusions are corroborated by an analysis of the

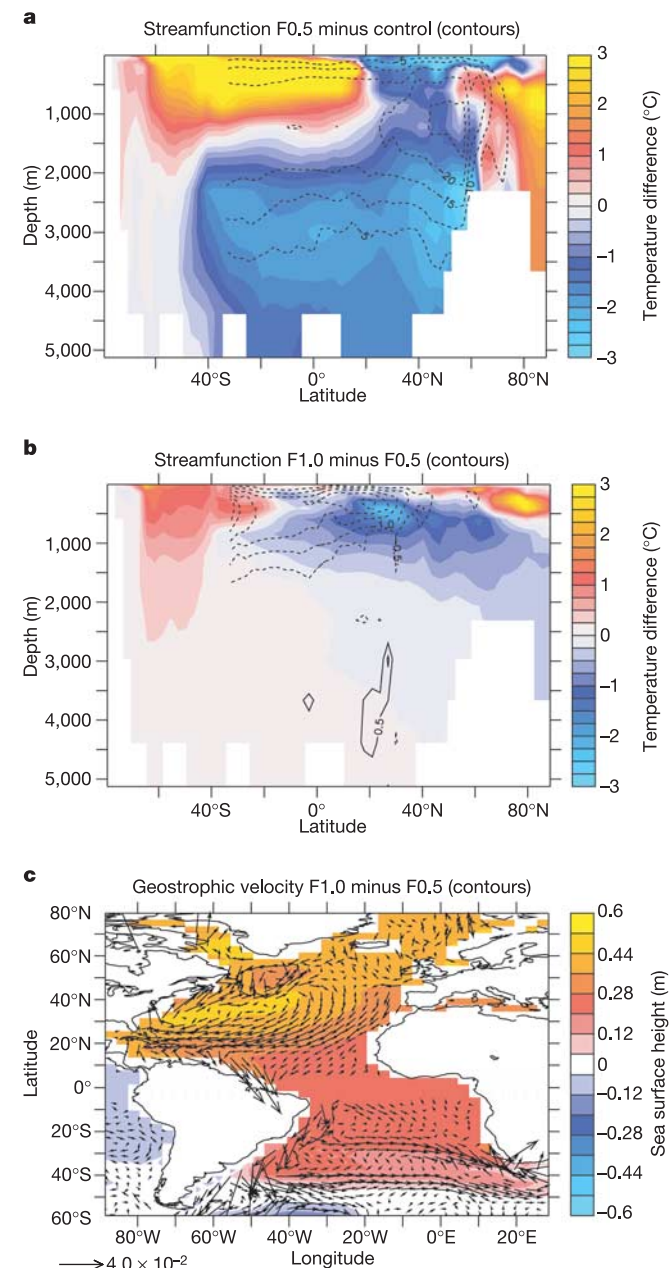


Figure 2 The ocean response to freshwater discharge into the North Atlantic. F1.0 and F0.5 denote 1.0 Sv and 0.5 Sv sustained freshwater discharge, respectively. **a**, The equilibrium Atlantic temperature difference (°C, colours) and meridional streamfunction difference (Sv, contours) resulting from 0.5 Sv freshwater input and the subsequent complete shutdown of the THC. **b**, Doubling the freshwater input leads to an additional temperature and streamfunction anomaly. **c**, Freshwater anomalies trigger a redistribution of the sea surface height (m, colours), owing to fast wave adjustment processes, which in turn drives geostrophic transport anomalies (m s^{-1} , vectors). Subsequently a cross-equatorial anomalous meridional transport is established (contours, **b**), which leads to an export of heat into the Southern Ocean of about 0.1 PW, thereby warming the upper 800 m of the Southern Ocean. On the other hand, associated surface temperature signals in the North Atlantic are weak.

thermal wind balance, as deduced from the upper-ocean zonal density gradients. The ocean adjustment scales linearly with the freshwater perturbation. The freshwater amplitudes in Fig. 2 are only chosen to be large to separate the related signal better from the internal ocean variability.

From these idealized simulations we conclude that the Southern Ocean surface temperature changes are not only determined by the heat transport of the large-scale THC, but are also directly affected by the anomalous meridional overturning circulation (Fig. 2b), which is established in response to the freshwater input in the North Atlantic. This provides a consistent explanation for the apparent but so far unexplained similarity between sea level variations and southern temperatures. Additional model simulations show that the same mechanism works equally well for freshwater input into the North Pacific, whereas it does not operate for freshwater input in the Southern Ocean, where zonal density anomalies cannot be established.

The thermal–freshwater seesaw

The direct effect of freshwater release on the meridional heat flux in the South Atlantic suggests a modification of the thermal bipolar seesaw concept²⁰. Assuming that the Southern Ocean acts like a heat reservoir, we propose the following approximation relating the freshwater flux F and the temperature anomalies T_N , T_S of the

northern and southern region relative to equilibrium climate:

$$d(T_S)/dt = (-aT_N + bF - T_S)/\tau \quad (1)$$

where τ is a typical thermal response timescale of the southern heat reservoir. In this concept, the southern reservoir temperature is controlled by the sum of a meridional oceanic heat transport due to the thermohaline circulation that is assumed to be proportional to $(-\alpha T_N - T_S)$ (ref. 20), and an oceanic heat transport that is proportional to $(\beta F - T_S)$, related to the freshwater input into the North Atlantic.

The three parameters in equation (1) are now determined by a series of transient simulations using the CGAOM (see Methods) in which the circulation is perturbed by a range of different freshwater discharges. The best fit yields a timescale τ of 114 yr; values for a and b are given in the Methods section. This timescale is considerably shorter than the one determined for the original thermal bipolar seesaw fitted to observations²⁰, which was of the order of 1,000 yr. In contrast, our new thermal–freshwater seesaw is based solely on the comprehensive CGAOM, and its agreement with palaeoclimatic proxy data provides an independent check of its validity. The extended thermal–freshwater seesaw concept is able to explain the timing and amplitude of the southern temperature response in the CGAOM in Fig. 1 with very high accuracy. We find that large changes in the freshwater discharge F contribute up to one-third to the amplitude of the southern temperature response T_S , and that the timescale τ is consistent with a thermal inertia timescale of the upper 1,000 m of the Southern Ocean.

An illustrative sequence of climate events

Forcing the CGAOM by freshwater discharge into the North Atlantic in a transient multi-millennial simulation reveals a picture of abrupt climate events that resolves a number of questions that emerged from earlier modelling studies. The illustrative freshwater scenario assumed in Fig. 3 causes large and abrupt changes in the North Atlantic deepwater formation and thus also in the air temperature of the North Atlantic region and Greenland with a magnitude of more than 15 °C in the convection regions. Changes in the modelled Antarctic temperature are up to about 5 °C and are controlled both by the temperature in the North Atlantic and by the amount of fresh water. Recent proxy estimates suggest changes in Antarctic temperature during the glacial of 2 to 4 °C (refs 25–27). This indicates that the peak of northern freshwater input mimicking a Heinrich surge, although intentionally chosen large here to illustrate its effect, is probably at the upper limit. The resulting sea level change of 30 m also suggests that the freshwater amplitude, although consistent with the largest estimate of sea level variations¹², is chosen rather high. The apparent lead of the peak Antarctic warming to the Greenland transition T2 depends on the shape of the freshwater discharge and would be smaller if fresh water was decreased more rapidly. Using the illustrative freshwater scenarios, the CGAOM simulates two types of events in the North Atlantic, a transition from a weak to a strong meridional overturning state (transition T1 in Fig. 3) and a transition from a stratified Atlantic without deepwater formation to a strong overturning state (transition T2). The spatial patterns (shown in Fig. 4) of temperature and precipitation changes of transition T2 suggest that climate changes following a Heinrich event were large in amplitude and seen in most regions of the globe. The qualitative model response is consistent with proxy reconstructions indicating broad cooling of the Southern Hemisphere and warming of the Northern Hemisphere during interstadial phases¹ and, for example, higher accumulation in Greenland, warmer and wetter conditions in Europe³⁷, wet conditions in most of northern subtropical Asia³⁸ and Arabia³⁹, as well as increased rainfall over the Cariaco basin⁴⁰. Many proxies distant from the North Atlantic record the large DO events following a Heinrich event (for example, DO events 8, 12) but not all of the smaller ones (for example, DO events 9, 13)⁴⁰. This is qualitatively

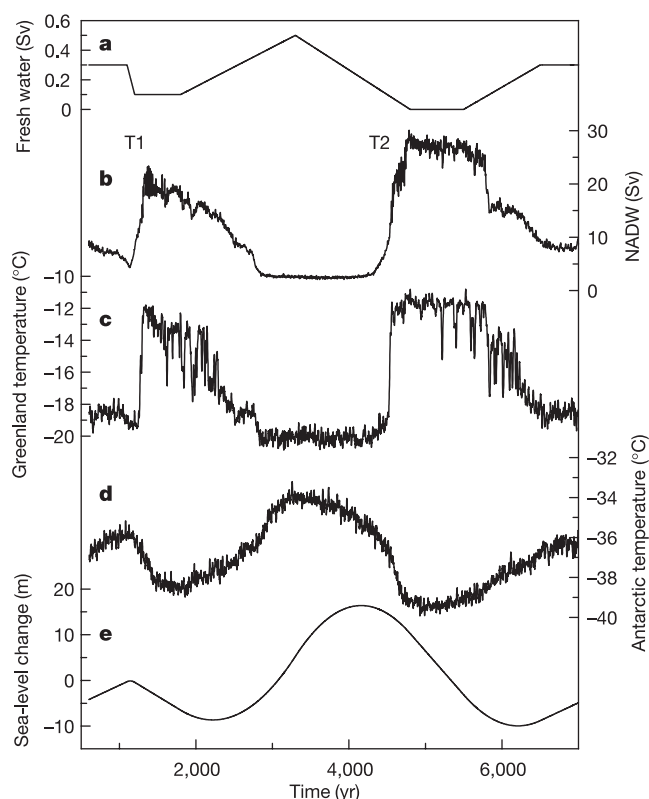


Figure 3 Time evolution of the THC and global sea level and corresponding changes in polar near-surface air temperature in an illustrative scenario of freshwater discharge into the North Atlantic. The fresh water (a) causes abrupt shifts in the North Atlantic deepwater formation (b). The associated massive and abrupt warming events simulated over Greenland (c) and the North Atlantic region are reminiscent of the DO events observed in palaeorecords from the last glacial epoch. Antarctic temperature (d) is influenced by the THC and directly by the freshwater discharge into the North Atlantic. This explains why changes in sea level (e, detrended integral of freshwater input) resemble but lag notably Antarctic temperature, consistent with the proxy reconstructions. Freshwater discharge in this scenario integrates to 30 m sea-level change, the upper limit of published values.

supported here by the smaller amplitude and spatial extent for transition T1 compared to T2. The modelled temperature shift in Greenland is similar for both transitions, also in good agreement with Greenland isotopic records²⁸, which indicate large temperature shifts even for the short DO events. The model response is in line with most proxies showing a hemispheric antiphase temperature pattern¹, and does not support in-phase changes in Greenland and Antarctica^{41,42}.

It is clear from Fig. 3 and from our thermal–freshwater seesaw that the shape of the sea level variations is similar to Antarctic temperature, as recently suggested from a reconstruction of sea level from the Red Sea¹². A significant amount of ice surges must therefore have emerged from the Northern Hemisphere, in line with the generally accepted pattern of the Heinrich surges recorded in the North Atlantic⁷. The results presented here do not preclude contributions to sea level from other locations⁴³. However, additional CGAOM simulations show that freshwater discharge into the North Pacific or into the Gulf of Mexico have a similar effect of warming Antarctica, whereas freshwater discharge from Antarctica has the opposite effect of cooling the Southern Ocean and Antarctica and warming the North Atlantic⁴⁴. Therefore, a dominant contribution from the Antarctic ice sheet to the sea level variations recorded during MIS 3 (refs 9–12) can be excluded.

We find that despite their similarity, sea level in the model lags Antarctic temperature by several centuries. Such a lag can be identified in palaeodata using the benthic oxygen isotope record from a sediment core off Portugal⁹, if taken as a proxy for sea level, and the Byrd oxygen isotope record²⁸ as a proxy for Antarctic temperature (see Fig. 5). Both records are independently synchro-

nized to the GRIP record from Greenland^{8,9}. Maximizing lag correlation in a moving window of 10 kyr suggests a lag of 300 to 1,500 yr of the benthic sea level curve to Antarctic temperature when using the published timescales. However, the uncertainties in the timescales (resolution of the records, gas age–ice age and synchronization uncertainties) and the fact that part of the benthic signal could be caused by changes in ocean temperature (which need not necessarily be in phase with sea level) prevent us from firmly concluding that there is a sea-level lag from the data alone. Equation (1) suggests almost no sea-level lag for changes in the freshwater flux F on short timescales (decades), but a phase lag of up to 90°, equivalent to about 1,000 yr, when assuming changes in F on timescales of a few millennia.

The sequence of events during MIS 3

The sequence of events shown in Fig. 3 is reminiscent of parts of the last glacial period, for example, the time around 45,000 yr before present (45 kyr before present, BP) with DO events 13 and 12. But the complexity of the CGAOM, the uncertainties in the hysteresis behaviour of the glacial thermohaline circulation and the computational cost of the model prevent us from simulating longer time periods. However, equation (1) has been shown to adequately relate the polar temperature anomalies simulated by the CGAOM and can thus be used as a substitute. Here we use the substitute conceptual model to quantify the extent to which the thermal–freshwater seesaw concept can explain the evolution of Greenland and Antarctic temperature reconstructions, as well as sea-level variations during Marine Isotope Stage (MIS) 3. We start from a random time evolution of the freshwater flux F . Subsequently, F is iteratively

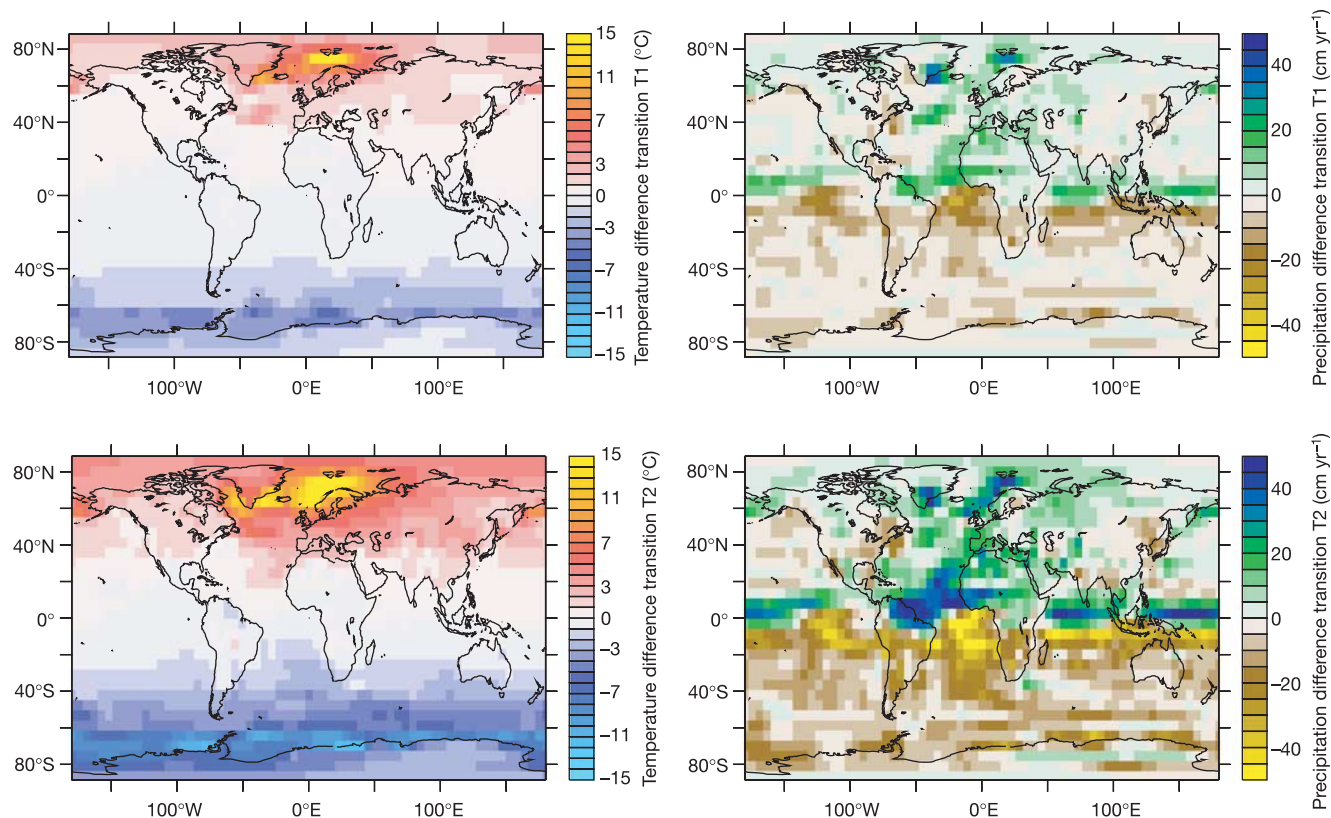


Figure 4 Temperature and precipitation changes simulated for two stadial–interstadial transitions. Transition T1 (shown in Fig. 3) is from a partial ‘off’ to a THC ‘on’ state, transition T2 from a completely collapsed THC state to a strong THC ‘on’ state. The CGAOM predicts that temperature changes were prominent in Greenland for both transitions, but the amplitude and spatial extent was different over most of the Northern Hemisphere. The maps show the near-surface temperature (°C) and precipitation

difference (cm yr^{-1}) for the cold-to-warm transition T1 (1,500 to 1,700 minus 800 to 1,000 model yr, top) and T2 (4,800 to 5,000 minus 3,400 to 3,600 model yr, bottom). These patterns are in agreement with proxy evidence for the long-lasting DO events following a Heinrich event (T2) and for the shorter DO events, which have a weak or no clear counterpart in the south (T1). The large temperature shifts in some areas of the Southern Ocean during transition T2 are caused by changes in sea-ice extent.

changed so as to maximize the correlation between the simulated time series T_N , T_S and sea-level h (obtained from the time integral of F) with the reconstructed time series GRIP $\delta^{18}\text{O}$, Byrd $\delta^{18}\text{O}$ and the sea-level proxy, respectively. T_N is assumed to be proportional to the strength of the THC. The latter is assumed to depend in a tanh way on the freshwater flux F , an approximation to a very narrow hysteresis. Thus, the expressions for the polar temperatures read $T_N = p \tanh(sF)$ and $d(T_S)/dt = (-aT_N + bF - T_S)/\tau$. Parameters are taken from the CGAOM (see Methods)^{25–27}. The optimized freshwater flux time series F and its relation to the reconstructed and simulated time series are depicted in Fig. 5.

Despite the simplicity of this approach, we find correlations of $r = 0.75$, 0.71 and 0.86 of the calculated curves with GRIP $\delta^{18}\text{O}$, Byrd $\delta^{18}\text{O}$ and the sea-level proxy, respectively. Therefore, this concept explains 60% (r^2) of the variability found in the polar temperature and sea-level reconstructions over MIS 3 (60 to 25 kyr BP). In particular, it shows that the slow changes in F explain the low-frequency timescale found in the Byrd isotopic record. Maximum values in the optimized freshwater flux coincide with high input of ice-rafted debris during Heinrich events⁷ and low values of benthic $\delta^{13}\text{C}$, indicating a reduced ventilation of the North Atlantic Deep Water⁹. Although we show here that both the shape of sea-level variations and marine proxies in the North Atlantic over MIS 3 are broadly consistent with meltwater discharge into the North Atlantic only, we note that this concept does not preclude additional meltwater discharge of smaller magnitude elsewhere. A change in the slope parameter s affects the amplitude needed in F

and can thus change the magnitude (but not the shape) of the sea-level variations by at least a factor of two, without significantly changing the correlation values. The magnitude of these sea-level variations in the last glacial epoch is still debated^{10–12}. The benthic record off Portugal⁹ probably reflects changes in both sea level and ocean temperature. A recent reconstruction of sea level¹² shows a very similar shape, which confirms the time evolution (but not necessarily the magnitude) of sea-level variations. However, this reconstruction does not have sufficient age control to be used in this study.

Substituting either the Byrd record or the Byrd and the sea-level record with red noise time series, the same concept can explain only about 40% (r^2) of the variability. This indicates that, in contrast to recent studies^{45,46}, we show here a strong and statistically significant ($>99\%$) coupling of the hemispheres during the climate variations in MIS 3.

Ocean linkage during abrupt climate change

Using a synthesis of proxy data from the last glacial period and a coupled climate model, we have demonstrated that the main patterns of temperature and sea-level variations recorded in a variety of archives during abrupt glacial climate events can largely be explained by changes in the oceanic heat transport related directly to freshwater discharge and the large-scale thermohaline circulation. This supports the view that the ocean circulation and its potential nonlinear changes play a crucial role in modulating the Earth's climate on global and regional scales. But there is also a strong response of the atmospheric circulation and of precipitation, indicating that atmospheric processes and feedbacks are relevant to transmit the abrupt Atlantic climate signals to other regions⁴⁷. We have obtained a quantitative picture of how the ocean circulation has shaped millennial-scale climate variability during the last glacial period, but are still unable to determine whether atmospheric processes, oceanic thresholds, ice-sheet dynamics, and their coupling or an external forcing acted as the pacemaker of abrupt climate changes. □

Methods

The model used for this study is the global coupled atmosphere–ocean–sea ice model ECBILT-CLIO (version 3.0). The atmosphere is represented by the T21, three-level quasi-geostrophic model ECBILT2 (ref. 29), which contains a full hydrological cycle and explicitly computes synoptic variability associated with weather patterns. The ocean model CLIO³⁰ is a primitive equation, free-surface ocean general circulation model with a resolution of 3×3 degrees and 20 unevenly spaced levels, coupled to a thermodynamic–dynamic sea-ice model. The CGAOM includes realistic topography, bathymetry, a simple representation of land surface processes and a bucket runoff scheme. Given the long timescales investigated here, the model is among the most complex climate models that can be applied to study this type of question at present. The model is freely available from <http://www.knmi.nl/onderzk/CKO/ecbilt.html>. The presented simulations use modern boundary conditions. Changing topography, greenhouse gas concentrations, orbital parameters and albedo values to conditions of the Last Glacial Maximum⁴⁸ show very similar results (see Supplementary Fig. S1). Freshwater discharge into the North Atlantic is between 50°N and 70°N , and is uniformly compensated outside the discharge regions to avoid model drift. The coupled model employs weak freshwater flux corrections. Given the large uncertainties in the freshwater budget of the Atlantic, the zero level in freshwater flux can thus be shifted within a wide range. Such model results are clearly sensitive to the shape of the hysteresis of the THC, which is mainly determined by the equilibrium freshwater budget of the Atlantic, and is often tuned by adjusting runoff masks or prescribing freshwater redistributions (flux corrections). Given the uncertainties associated with the THC hysteresis and the feedbacks, which determine the stability of the THC even under modern boundary conditions⁴⁹, we have chosen not to tune the CGAOM to achieve a certain threshold behaviour. Although the processes studied require a comprehensive general circulation ocean model, our conclusions do not depend on a particular threshold behaviour of the THC.

The parameters a , b and τ of equation (1) were determined by minimizing the sum of the squared deviations for southern temperature of the CGAOM and of equation (1), taking fresh water and northern temperature as an input. Near-surface temperatures of the North Atlantic and Southern Ocean region yield $a = 0.41^\circ\text{C}^\circ\text{C}^{-1}$, $b = 3.6^\circ\text{C}\text{Sv}^{-1}$ and $\tau = 114\text{ yr}$ (used for Fig. 1). Correlation of the CGAOM results (10-yr running means) and the response of equation (1) with optimal parameters exceeds $r = 0.97$ in all cases. For the original thermal bipolar seesaw (equation (1), $b = 0$), the fitting procedure yields $\tau = 90\text{ yr}$.

For the conceptual model used in Fig. 5, all parameters except the slope s are fitted from

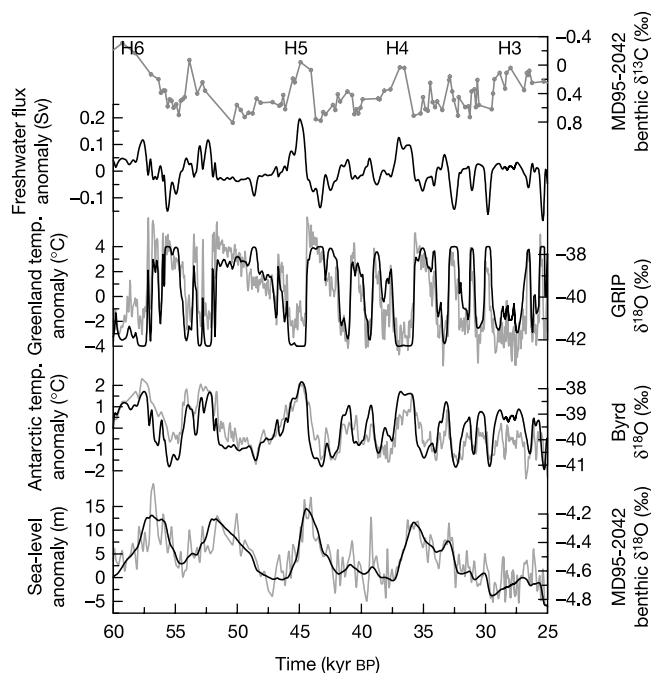


Figure 5 Variability of Greenland and Antarctic temperature and sea-level proxy data (grey, right axes) explained by a conceptual model (black, left axes). The optimized solution to the conceptual model (see text and Methods) shows strong coupling of Greenland and Antarctic temperature, fresh water and sea level during MIS 3 and explains 60% (r^2) of the variability seen in the water isotope reconstructions of GRIP², Byrd²⁸ and the benthic isotope record from off Portugal⁹, which is assumed to be a proxy for sea-level variations. The timing of the optimized freshwater discharge peaks shows remarkable agreement with input of ice-rafted debris during Heinrich events⁷ and low values of benthic $\delta^{13}\text{C}$, indicating reduced ventilation of the North Atlantic Deep Water⁹ (grey, top). Age scales are synchronized to GRIP^{9,9} and all time series except $\delta^{13}\text{C}$ are splined with a 100-yr cut-off period⁵⁰. Note that benthic $\delta^{13}\text{C}$ and ice-rafted debris are not used to optimize correlation.

the Greenland and Antarctic temperature response of the CGAOM and yield $a = 0.24^{\circ}\text{C}^{\circ}\text{C}^{-1}$, $b = 6.3^{\circ}\text{C Sv}^{-1}$, $\tau = 135\text{ yr}$ and $p = 4^{\circ}\text{C}$. The range of Antarctic temperature variations is required to be in agreement with observations^{25–27}. Models simulating the sensitivity of the THC to fresh water indicate that the range for the slope s is uncertain. However, the choice of the value s is not critical for our conclusion. That only one branch of the THC hysteresis behaviour is captured in our simple model explains why the calculated temperature generally drops too fast at the end of the Antarctic warm events, as compared to reconstructions. Ocean models indicate that the lower branch of the hysteresis ('off-on' transition) is probably more abrupt than the upper one. Such a hysteresis behaviour would need a slower decrease in the freshwater flux and thus lead to a slower decrease in Antarctic temperature, while still producing an abrupt warming in Greenland, and thus improve the agreement with proxy data.

Received 1 March; accepted 25 June 2004; doi:10.1038/nature02786.

1. Voelker, A. H. L. & workshop participants. Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: A database. *Quat. Sci. Rev.* **21**, 1185–1212 (2002).
2. Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250 kyr ice-core record. *Nature* **364**, 218–220 (1993).
3. Schwander, J. *et al.* Age scale of the air in the summit ice: Implication for glacial-interglacial temperature change. *J. Geophys. Res.* **102**, 19483–19494 (1997).
4. Lang, C., Leuenberger, M., Schwander, J. & Johnsen, S. 16°C rapid temperature variation in central Greenland 70,000 years ago. *Science* **286**, 934–937 (1999).
5. Severinghaus, J. P. & Brook, E. J. Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice. *Science* **286**, 930–934 (1999).
6. Heinrich, H. Origin and consequences of cyclic ice rafting in the northeast Atlantic Ocean during the past 130,000 years. *Quat. Res.* **29**, 142–152 (1988).
7. Hemming, S. R. Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42** (2004) doi:10.1029/2003RG000128.
8. Blunier, T. & Brook, E. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
9. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial scale events 64,000 to 24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
10. Yokoyama, Y., Esat, T. M. & Lambeck, K. Coupled climate and sea-level changes deduced from Huon Peninsula coral terraces of the last ice age. *Earth Planet. Sci. Lett.* **193**, 579–587 (2001).
11. Chappell, J. Sea level changes forced ice breakouts in the last glacial cycle: New results from coral terraces. *Quat. Sci. Rev.* **21**, 1229–1240 (2002).
12. Siddall, M. *et al.* Sea-level fluctuation during the last glacial cycle. *Nature* **423**, 853–858 (2003).
13. Clark, P. U., Webb, R. S. & Keigwin, L. D. (eds) *Mechanisms of Global Climate Change at Millennial Time Scales* 1–394 (AGU, Washington DC, 1999).
14. Stocker, T. F. & Marchal, O. Abrupt climate change in the computer: Is it real? *Proc. Natl Acad. Sci. USA* **97**, 1362–1365 (2000).
15. Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
16. Crowley, T. J. North Atlantic deep water cools the southern hemisphere. *Paleoceanography* **7**, 489–497 (1992).
17. Stocker, T. F. The seesaw effect. *Science* **282**, 61–62 (1998).
18. Broecker, W. S. Paleocene circulation during the last deglaciation: A bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
19. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
20. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, doi:10.1029/2003PA000920 (2003).
21. Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
22. Mikolajewicz, U. & Maier-Reimer, E. Mixed boundary conditions in ocean general-circulation models and their influence on the stability of the models conveyor belt. *J. Geophys. Res.* **99**, 22633–22644 (1994).
23. Rahmstorf, S. Rapid climate transitions in a coupled ocean-atmosphere model. *Nature* **372**, 82–85 (1994).
24. Schmittner, A., Yoshimori, M. & Weaver, A. J. Instability of glacial climate in a model of the ocean-atmosphere-cryosphere system. *Science* **295**, 1489–1493 (2002).
25. Jouzel, J. *et al.* Magnitude of isotope/temperature scaling for interpretation of central Antarctic ice cores. *J. Geophys. Res.* **108**, doi:10.1029/2003JD002677 (2003).
26. Blunier, T., Schwander, J., Chappellaz, J., Parrenin, F. & Barnola, J.-M. What was the surface temperature in central Antarctica during the last glacial maximum? *Earth Planet. Sci. Lett.* **218**, 379–388 (2004).

27. Stenni, B. *et al.* A late-glacial high-resolution site and source temperature record derived from the EPICA Dome C isotope records (East Antarctica). *Earth Planet. Sci. Lett.* **217**, 183–195 (2003).
28. Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. Jr Oxygen isotope profiles through the Antarctic and Greenland ice sheets. *Nature* **235**, 429–434 (1972).
29. Opsteegh, J. D., Haarsma, R. J., Seltin, F. M. & Kattenberg, A. ECBILT: A dynamic alternative to mixed boundary conditions in ocean models. *Tellus A* **50**, 348–367 (1998).
30. Goosse, H. & Fichefet, T. Importance of ice-ocean interactions for the global ocean circulation: A model study. *J. Geophys. Res.* **104**, 23337–23355 (1999).
31. Stocker, T. F., Wright, D. G. & Broecker, W. S. The influence of high-latitude surface forcing on the global thermohaline circulation. *Paleoceanography* **7**, 529–541 (1992).
32. Schiller, A., Mikolajewicz, U. & Voss, R. The stability of the thermohaline circulation in a coupled ocean-atmosphere general circulation model. *Clim. Dyn.* **13**, 325–348 (1997).
33. Manabe, S. & Stouffer, R. J. Coupled ocean-atmosphere model response to freshwater input: Comparison to Younger Dryas event. *Paleoceanography* **7**, 321–336 (1992).
34. Huang, R. X., Cane, M. A., Naik, N. & Goodman, P. Global adjustment of the thermocline in response to deepwater formation. *Geophys. Res. Lett.* **27**, 759–762 (2000).
35. Hsieh, W., Davey, M. K. & Wajsiwicz, C. The free Kelvin wave in finite-difference models. *J. Phys. Oceanogr.* **13**, 1383–1397 (1983).
36. Johnson, H. & Marshall, D. P. A theory for the surface Atlantic response to thermohaline variability. *J. Phys. Oceanogr.* **32**, 1121–1132 (2002).
37. Allen, J. R. M. *et al.* Rapid environmental changes in southern Europe during the last glacial period. *Nature* **400**, 740–743 (1999).
38. Wang, Y. J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
39. Burns, S. J., Fleitmann, D., Matter, A., Kramers, J. & Al-Subbary, A. A. Indian ocean climate and an absolute chronology over Dansgaard/Oeschger events 9 to 13. *Science* **301**, 1365–1367 (2003).
40. Peterson, L. C., Haug, G. H., Hughen, K. A. & Röhl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last glacial. *Science* **290**, 1947–1951 (2000).
41. Steig, E. J. *et al.* Synchronous climate changes in Antarctica and the North Atlantic. *Science* **282**, 92–95 (1998).
42. Mulvaney, R. *et al.* The transition from the last glacial period in inland and near-coastal Antarctica. *Geophys. Res. Lett.* **27**, 2673–2676 (2000).
43. Kanfoush, S. L. *et al.* Millennial-scale instability of the Antarctic Ice Sheet during the last glaciation. *Science* **288**, 1815–1818 (2000).
44. Weaver, A. J., Saenko, O. A., Clark, P. U. & Mitrovica, J. X. Meltwater pulse 1A from Antarctica as a trigger of the Bølling-Allerød warm interval. *Science* **299**, 1709–1713 (2003).
45. Wunsch, C. Greenland-Antarctic phase relations and millennial time-scale climate fluctuations in the Greenland ice-cores. *Quat. Sci. Rev.* **22**, 1631–1646 (2003).
46. Roe, G. H. & Steig, E. J. Characterization of millennial-scale climate variability. *J. Clim.* **17**, 1929–1944 (2004).
47. Broecker, W. S. Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO₂ upset the current balance? *Science* **278**, 1582–1588 (1997).
48. Timmermann, A., Justino, F. B. & Jin, F.-F. Surface temperature control in the north and tropical Pacific during the last glacial maximum. *Clim. Dyn.* (in the press).
49. IPCC. *Climate Change: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge/New York, 2001).
50. Enting, I. G. On the use of smoothing splines to filter CO₂ data. *J. Geophys. Res.* **92**, 10977–10984 (1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the model developers and KNMI for making ECBILT-CLIO available to the scientific community, to F. Justino and U. Krebs for setting-up the glacial version of the model, and to N. Shackleton for discussions. This work was supported by the Swiss National Science Foundation, the Swiss Federal Office of Science and Education through the EC project POP and the University of Bern. A.T. was supported by the Deutsche Forschungsgemeinschaft through a Collaborative Research Project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (knutti@climate.unibe.ch).

The structure and evolution of centromeric transition regions within the human genome

Xinwei She^{1,2*}, Julie E. Horvath^{1*}, Zhaoshi Jiang^{1,2}, Ge Liu¹, Terrence S. Furey³, Laurie Christ¹, Royden Clark¹, Tina Graves⁴, Cassy L. Gulden¹, Can Alkan¹, Jeff A. Bailey¹, Cenik Sahinalp^{1,5}, Mariano Rocchi⁶, David Haussler³, Richard K. Wilson⁴, Webb Miller⁷, Stuart Schwartz¹ & Evan E. Eichler^{1,2}

¹Department of Genetics, Center for Computational Genomics and the Center for Human Genetics, Case Western Reserve University School of Medicine and University Hospitals of Cleveland, Cleveland, Ohio 44106, USA

²Department of Genome Sciences, University of Washington School of Medicine, 1705 NE Pacific St, Seattle, Washington 98195, USA

³UCSC Genome Bioinformatics Group, Center for Biomolecular Science & Engineering, University of California, Santa Cruz, 1156 High St, Santa Cruz, California 95064, USA

⁴Washington University School of Medicine, Genome Sequencing Center, 4444 Forest Park Boulevard, St Louis, Missouri 63108, USA

⁵School of Computing Science, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada

⁶Sezione di Genetica, DAPEG, University of Bari, Via Amendola 165/A 70126 Bari, Italy

⁷Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

* These authors contributed equally to this work

An understanding of how centromeric transition regions are organized is a critical aspect of chromosome structure and function; however, the sequence context of these regions has been difficult to resolve on the basis of the draft genome sequence. We present a detailed analysis of the structure and assembly of all human pericentromeric regions (5 megabases). Most chromosome arms (35 out of 43) show a gradient of dwindling transcriptional diversity accompanied by an increasing number of interchromosomal duplications in proximity to the centromere. At least 30% of the centromeric transition region structure originates from euchromatic gene-containing segments of DNA that were duplicatively transposed towards pericentromeric regions at a rate of six–seven events per million years during primate evolution. This process has led to the formation of a minimum of 28 new transcripts by exon exaptation and exon shuffling, many of which are primarily expressed in the testis. The distribution of these duplicated segments is nonrandom among pericentromeric regions, suggesting that some regions have served as preferential acceptors of euchromatic DNA.

Centromeres and the corresponding euchromatic transition regions have been described as one of the last frontiers of eukaryotic genome sequencing¹. Before the sequencing of the human genome, the model for the organization of these regions was relatively simple (Fig. 1a)^{2,3}. Although a more complex organization has begun to become apparent (Fig. 1b), the true sequence nature of these transition regions remained poorly characterized during the initial draft of the human genome, due in part to the paucity of unique mapping reagents near centromeres and artefacts associated with sequence and assembly of duplicated sequences^{4–6}. During the final phases of the ‘completion’ of the human genome, considerable resources were directed towards improving the quality of these problematic areas (see Methods and Supplementary Methods).

Centromeric transition regions and assembly quality

We assessed the completeness of centromeric transition regions within the finished genome (build34, July 2003) by a series of experimental and computational analyses (Fig. 2; see also Supplementary Tables 1–4). First, we analysed the sequence composition of each of the 43 targeted pericentromeric regions. We found that 29 out of 43 (67.4%) of these show a minimum of 10 kilobases (kb) of satellite sequence positioned within the most proximal location of each chromosome arm (Fig. 2). Seven show a near-perfect match with higher-order alpha-satellite DNA (Supplementary Table 2 and Supplementary Methods). These proximal sequence features are consistent with centromere DNA structure (see Fig. 1). As expected, an abundance of duplicated segments is

observed in close proximity to alpha-satellite DNA (<5 Mb). Gaps in the sequence assembly are particularly prevalent in these areas and show the strongest association with segmental duplications (76 out of 78 pericentromeric gaps are flanked by segmental duplications). Using a fluorescence *in situ* hybridization (FISH)-based assay to assess the multi-site distribution of segmental duplications, we estimate that 26.7% (82 out of 307 signals) of pericentromeric duplications are absent (Methods; see also Supplementary Tables 5, 6). A second, sequence-based assay calculates that approximately 34% of sequence-tagged sites cannot be identified within the current finished genome assembly (Supplementary Table 7). We conservatively estimate that ~4 Mb of satellite-rich sequence and ~6.5 Mb of highly duplicated material remain to be sequenced as part of these transition regions. This is in addition to the estimated 200 Mb of missing sequence that constitutes heterochromatic and acrocentric portions of the human genome.

It should be noted that this clone-order-based assembly (build 34) differs significantly from whole-genome shotgun sequence assembly (WGSA) of the human genome⁷. We analysed a recently published WGSA of the human genome and found that in this assembly an additional 19% (38.2 Mb) of the pericentromeric sequence is not assembled (24 Mb), not assigned (11.3 Mb) or misassigned (2.3 Mb). We estimate that more than 40% of the duplicated sequence presently assembled within build 34 might be incorrectly mapped by this WGSA assembly. The clone-order-based assembly of the human genome, therefore, provides one of the first detailed views of the organization of centromeric transition regions within mammalian genomes. As most of the eukaryotic genome projects have now

adopted WGS strategies, it is unlikely that such regions will be readily resolved in the future unless targeted efforts are undertaken.

Duplication organization

We assessed a variety of sequence properties (repeat content, duplication features, exon density, per cent G + C composition, and so on) as a function of distance from each putative centromere (Fig. 3; see also Supplementary Fig. 1a, b). Two significant features were noted: reduced gene density and increased duplication content. On the basis of our global analysis of segmental duplications (E.E.E., manuscript in preparation)^{6,8}, we have developed a working model for human pericentromeric organization (Fig. 1b). The majority of human pericentromeric DNA (29 out of 43 chromosome arms) shows evidence of blocks of segmental duplication located in close proximity to centromeric satellite DNA. Within a 5-Mb window of the centromere, we found that 22.7% (47.2 out of 207.9 Mb) of the bases are duplicated (sequence identity and length thresholds of >90% and >1 kb, respectively). Pericentromeric regions account for 31.1% of all duplicated bases (47.2 out of 152 Mb) and nearly 33.2% of all pairwise alignments (8,384 out of 25,239) for the entire human genome. Simulations confirm that segmental duplications are significantly ($P < 0.0001$) enriched (six–sevenfold) near centromeres when compared with a random genome model.

The pericentromeric enrichment for segmental duplications is most markedly seen for interchromosomal pairwise alignments where 5,357 out of 14,860 (36.0%) of all duplications between chromosomes occur within the first 5 Mb of the centromere. The proportion of interchromosomal duplications is most pronounced the closer the proximity to the centromere, where a clear gradient effect is observed (Fig. 3). Within the first 500 kb, interchromosomal duplications outnumber intrachromosomal duplications 6 to 1.

The proportion of interchromosomal to intrachromosomal pairwise alignments drops from 2.6 to 1.66 as 2 Mb and 5 Mb pericentromeric regions are considered, respectively. By 4.5 Mb, there is a noticeable decline in all duplications. Centromeric satellite sequences were significantly ($P < 0.001$) enriched precisely at the integration sites of segmental duplication, suggesting that satellite sequences have had a role in this process of non-homologous interchromosomal exchange⁹. In addition to these sequences, various classes of low complexity and simple repeat sequences mapped within 500 base pairs (bp) of the duplication boundaries (Supplementary Table 8). We observed a correlation ($r^2 = 0.4509$) between the number of such repeat elements and the number of pericentromeric duplications.

Although pericentromeric regions are, in general, enriched for duplication, there is considerable variability. Three groups may be distinguished. Pericentromeric regions where the duplication content is below the genome average (<5.2%) (5p11, 4q11, 19q11, 18q11, 8p11, Xp11, 6q11 and 16q11) show a relatively sharp transition between unique and alpha-satellite DNA^{10–12}. Sixteen pericentromeric regions (1p11, 3p11, 3q11, Xp11, 4p11, 5q11, 8q11, 17q11, 12p11, 19p11, 11q11, 12q11, 14q11, 20q11, 20p11 and Yq11) show an intermediate level of duplication between the genome and pericentromeric average (5.2–32.2%). Nineteen pericentromeric regions (10q11, 16p11, Yp11, 13q11, 2q11, 6p11, 7q11, 11p11, 21q11, 22q11, 10p11, 18p11, 17p11, 7p11, 1q11–1q12, 2p11, 15q11, 9q11 and 9p11) show extensive zones of duplication ranging from 500 kb to 5.5 Mb in length (Fig. 2, see also Supplementary Table 1c–f).

Specific constellations of pericentromeric regions share a greater number of, and in general have larger, segmental duplications. Regions 16p11, 15q11, 2p11, 7p11, 7q11 and 22q11 define one of the largest cohorts with approximately 22.7% (10.7 out of 47.2 Mb)

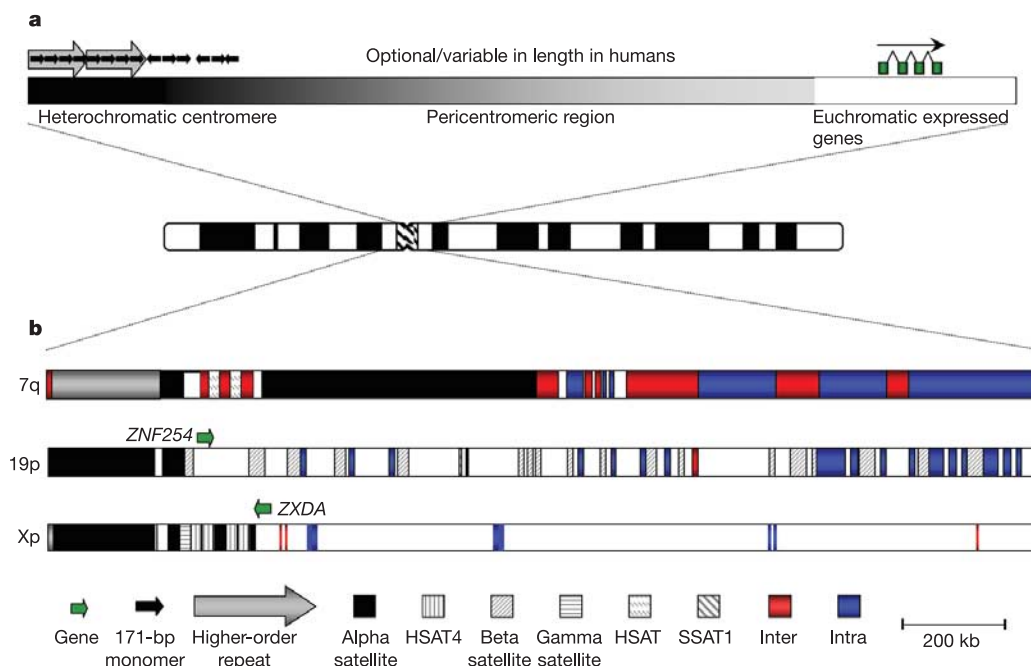


Figure 1 Models of centromeric transition regions. **a**, Pre-genome sequence model of pericentromeric organization: tandem reiterations of higher-order alpha-satellite DNA constitute larger array structures whose precise composition is diagnostic for a particular chromosome³⁵. Blocks of alpha-satellite DNA lacking higher-order structure as well as other pericentromeric satellite DNA sequences map to the periphery^{2,11,36,37}. In some cases, such as 9q12, 16q12 and 1q12, these peripheral satellite DNAs became

sufficiently large to warrant their own cytological designations known as a secondary constriction^{36–38}. **b**, Models of pericentromeric organization based on three sequenced chromosomes^{11,26,39} showing various degrees of duplication content and interstitial satellite content⁴⁰. Chromosome 7q represents a high level of segmental duplication whereas 19p represents an intermediate level and Xp a low level of segmental duplication.

of all duplicated bases shared between these six regions (Fig. 4; see also Supplementary Fig. 2). 18p11/21q11 and 9p11/20q11 define two smaller, more ancient, associations (Fig. 4). This distinction between quiescent and active regions of pericentromeric duplication is generally supported by our detailed FISH analyses of pericentromeric regions (Supplementary Tables 5, 6) and is, therefore, unlikely to be an artefact of missing sequence. We cannot exclude the possibility, however, that such regions may have become active in different primate lineages, as our analysis is based largely on examination of the contemporary human genome structure. In addition to the 43 pericentromeric regions near active centromeres, three other ancestral centromeres (2q21, 9q12 and 15q25.6) have been described^{13–15} that were inactivated by chromosomal rearrangements during the human lineage of evolution. Each of these regions was marked by an abundance of pericentromeric duplications (Supplementary Table 1c, d). As suggested earlier¹³, euchromatic regions characterized by an abundance of pericentromeric duplications may provide an evolutionary footprint of ancestral primate centromeres that have subsequently disappeared as a result of chromosomal fusion events. We identified five additional regions of the human genome (3p12, 3q21, 7q11, 13q12 and 22q11) that were marked by the presence of satellite sequences and an abundance of pericentromeric duplications.

Our detection of segmental duplications is based on arbitrary sequence identity and length thresholds (>90% and >1 kb). To address the question of whether the paucity of segmental duplications for certain chromosomes might simply be a consequence of these criteria, we performed a genome-wide analysis to detect more divergent segmental duplications (>250 bp and >75% sequence identity) (Supplementary Fig. 3). For those pericentromeric regions that harbour extensive duplication, the transition region between unique and duplicated regions has remained relatively precise even when the placement of more divergent duplications is considered. Among the duplication-quiescent centromeres (regions 5p11, 3q11, 4q11, 18q11 and 6q11) virtually no additional duplications were detected near the centromere, indicating that these regions have not been targets of segmental duplication during the entire course of human-primate evolution (Supplementary Fig. 3). Notably, for 8p11, 16q11 and 19q11, we observed small patches (<200 kb) of segmental duplication (<90% identity) that extended distally from the centromere. One interpretation may be that these regions were once capable of accepting duplications but subsequently became quiescent during the last 40 million years of chromosome evolution, as the segmental duplications all show >10% sequence divergence¹⁶. Another explanation for quiescent pericentromeric regions may be that chromosome rearrangement¹⁷ or centromere-repositioning events have uncoupled active centromeres and zones of pericentromeric duplication such that certain regions now appear quiescent. The pericentric inversion of 18q11 specifically within the human lineage¹⁴ and the recent emergence of the chromosome 6 centromere¹⁸ might explain the dearth of pericentromeric duplications for these chromosome arms.

Euchromatic colonization of human pericentromeric DNA

To understand further the evolutionary dynamics of pericentromeric DNA, we targeted one chromosome, 2p11, for a more systematic analysis. We constructed the first sequence contig (737 kb) representing an autosomal transition from euchromatin to higher-order alpha-satellite DNA (Fig. 5). We validated the organization of the region by Southern, paralogous sequence-tagged site (STS) content and extended fibre-FISH analysis (Supplementary Methods and J.E.H., unpublished data). It should be pointed out that sequence closure in this region required extensive sequence redundancy as well as considerably more validation than other relatively 'unique' regions of the human genome owing to the presence of highly identical duplications as well as large-scale structural variation among different chromosomal haplotypes.

Our initial sequence analysis showed that 91% of the 737-kb region consisted of segmental duplications.

Two types of duplication alignment were distinguished within 2p11. Fifty-seven per cent of the duplicated bases mapped to interstitial euchromatin located outside pericentromeric regions (Methods) whereas 34% of the duplications were part of alignments that mapped exclusively between pericentromeric regions. We classified these as ancestral duplicons and pericentromeric interspersed repeats, respectively^{12,19}. Only 2% of the 2p11 sequence was devoid of duplications whereas the remaining alignments (~7%) mapped between pericentromeric and subtelomeric regions or the Y chromosome. Both the ancestral duplicons and pericentromeric repeats were distributed among multiple pericentromeric regions, most often as part of larger pairwise alignments. The pattern of duplication indicated a mosaic organization that had been formed by the duplicative transposition of at least 13 different ancestral euchromatic regions followed by secondary rounds of pericentromeric duplication^{12,20–22}. To validate this euchromatic origin of the ancestral duplicons, we investigated nine of these regions in more detail. STS were designed within each duplicon and hybridized to genomic libraries from orang-utan (CHORI-251) and baboon (RPCI-41). All positive non-human primate clones were end-sequenced and mapped to the human genome by sequence similarity searches. In 7 out of 9 cases, the baboon data were consistent with a single, non-duplicated locus that mapped to a non-pericentromeric region of the human genome (Supplementary Table 9). In 6 out of 9 cases, the orang-utan clones mapped to the same locus, suggesting that these sites had been distributed to the pericentromeric region of human chromosomes relatively recently during evolution (<14 million years ago). Finally, mouse-human synteny analysis showed collinearity of syntenic anchors extending from the duplicated region into genomic sequence that flanked the putative ancestral segment but was not duplicated (Supplementary Fig. 4). These results unambiguously confirmed these sites as ancestral donor regions and provided directionality to the duplication events. An analysis of 15q11 similarly identified at least 16 different ancestrally donated euchromatic regions (Supplementary Fig. 5).

On the basis of our analysis of 2p11 and 15q11 (Locke, D. P. *et al.*, manuscript in preparation) as well as detailed studies of other human pericentromeric regions^{8,12,20–28}, we sought to identify the ancestral origin of all pericentromeric duplications that had emerged within the last 35–40 million years of human evolution (<10% sequence divergence). An ancestral locus was considered if it met three criteria: (1) it is not located within 5 Mb of the centromeric DNA; (2) most of the pairwise alignments underlying the locus map to pericentromeric regions of the human genome; and (3) mouse conserved synteny extends beyond the duplication alignment as determined by BLASTZ comparisons (<http://www.genome.ucsc.edu>) (Supplementary Methods). We analysed 8,343 pericentromeric duplication alignments and identified 271 (741 pairwise alignments) regions that met these criteria (Supplementary Table 10). These ancestral duplications correspond to 29.4% of all pericentromeric duplications (13.9 out of 47.2 Mb) within a 5-Mb window of the centromere. The putative ancestral donor loci ranged in length from 1 kb to 586 kb (average = 39.4 kb; median = 9.1 kb) and were duplicated on average to 2.73 (741/271) different pericentromeric regions. A total of 109 out of 271 (40%) of these donor sites contained intron–exon structure, suggesting that this process had been responsible for the mobilization of entire genes or partial gene fragments.

These data indicate that at least 30% of human pericentromeric duplications originated as transposed euchromatic sequence that was dispersed towards centromeric regions during hominoid chromosome evolution. By count or total number of duplicated bases, most of the ancestral duplications showed 94–97% sequence identity (Supplementary Fig. 6). We observed a marked reduction (twofold by count and tenfold by number of duplicated bases) for

ancestral duplications that showed >98% sequence identity. Comparative and phylogenetic data suggest a continuum of events with a particular burst of activity after the separation of the Old World monkey species but before the radiation of the great-ape species (0.02 to 0.05 substitutions per site; 10–25 million years ago). Subsequent pericentromeric–pericentromeric duplications differentially distributed blocks within specific great-ape lineages leading to quantitative and qualitative differences^{8,12,20–22,29}.

Whereas donor loci appear to be randomly distributed, the pericentromeric dispersal was not uniform (Table 1). Several pericentromeric regions are significantly enriched ($P < 0.0012$), indicating that these particular regions have been preferential acceptors of duplicatively transposed material whereas others may have been protected or may appear quiescent owing to recent large-scale deletion or rearrangement. Four pericentromeric regions alone (7q11, 16p11, 15q11, and 17p11) account for 37.4% of the ancestral

duplication alignments (277 out of 741) (Table 1). The extent of pericentromeric duplications among other non-human primates and mammalian organisms has only begun to be addressed^{30–32}. Our analysis, however, predicts that the current sequence architecture of many human centromeric transition regions is a derived property where syntenic relationships rapidly decay.

Pericentromeric transcripts

For most species, pericentromeric regions are generally regarded as transcriptionally poor^{1,33}. We measured transcript and gene density as a function of distance from each centromere using annotated known genes, Refseq genes and spliced expressed sequence tags (ESTs) (Fig. 3b; see also Supplementary Table 11a, b). Gene and exon density gradually increase as distance from the centromere increases. A noticeable reduction in exon density was observed within 2 Mb of the centromere when compared to the genome

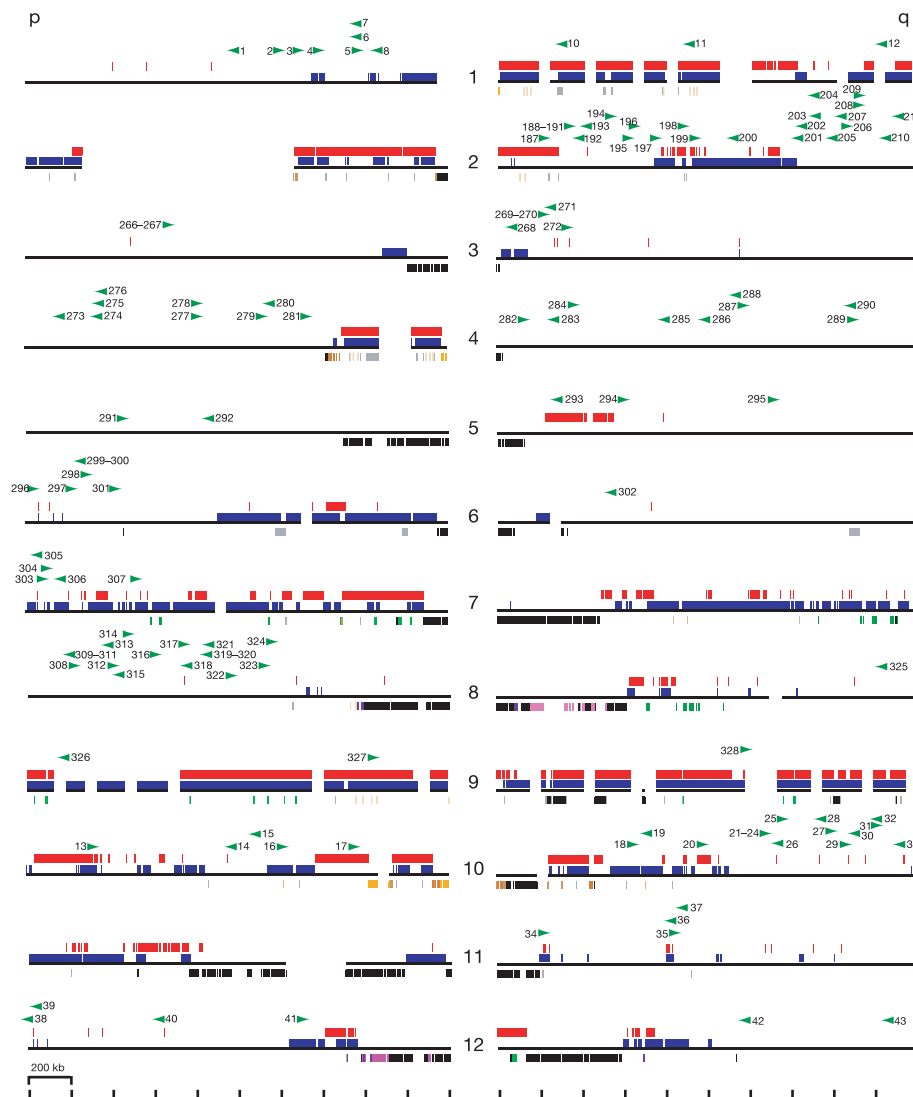


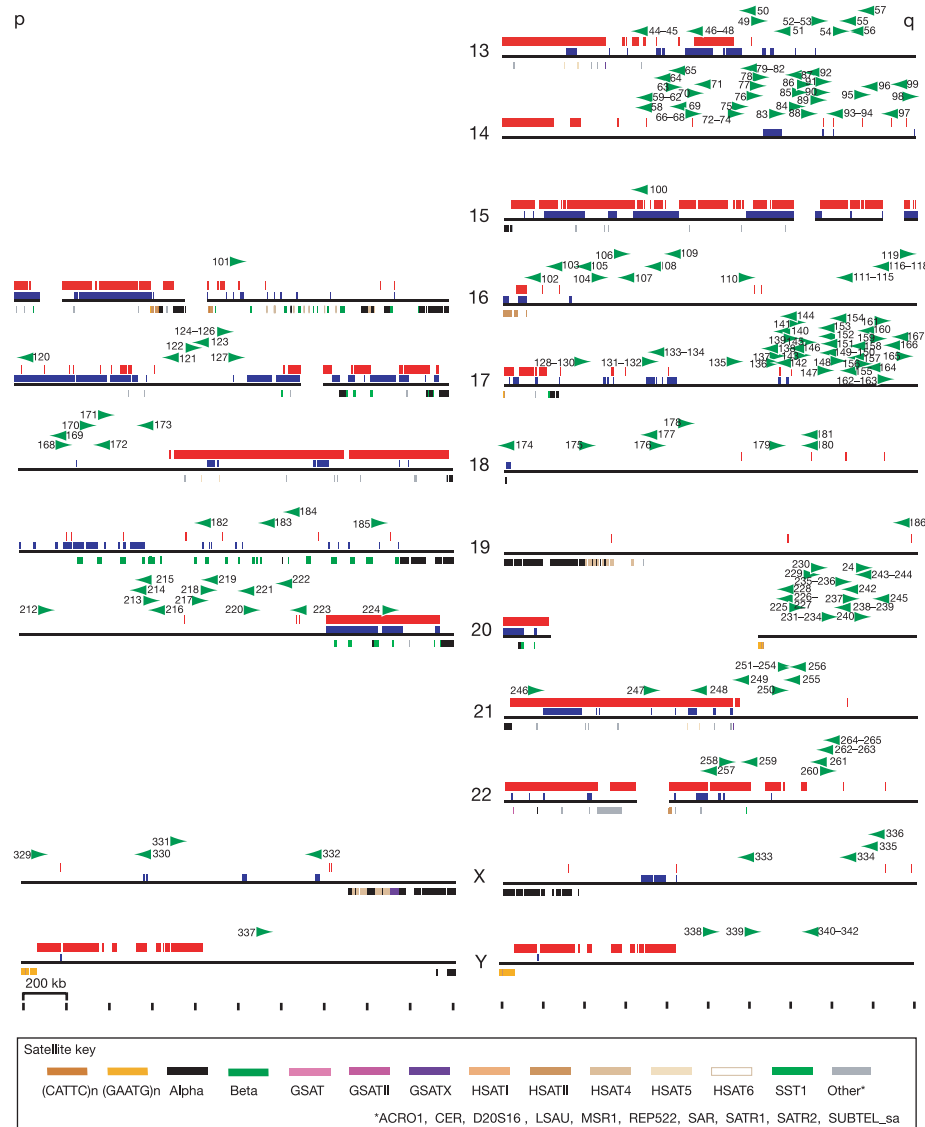
Figure 2 Pericentromeric architecture. The first 2 Mb on either side (p and q) of the centromere are shown for each chromosome. The centromere position was defined as the most proximal base pair for each chromosome arm based on the finished genome assembly (July 2003). Intrachromosomal (blue) and interchromosomal (red) duplications (>90% sequence identity, >1 kb in size) are shown above the line. Gaps (>50 kb in size) are shown as breaks in the black line. Centromeric satellite repeat composition is shown below the horizontal line according to the key. Significant blocks (>10 kb) of alpha-satellite or other centromeric satellite DNA is observed for 29 out of 43 pericentromeric regions. Acrocentric arms were not targeted as part of the human genome project and are therefore not shown. See Supplementary Fig. 3 for the distribution of pairwise alignments over a larger pericentromeric region (10 Mb), as well as a wider range of alignment divergence (0–0.25 substitutions per bp). Approximate locations of known genes are depicted by green arrows indicating the direction of transcription (see Supplementary Table 10a for corresponding gene names).

average (4.22 genes per Mb versus 7.3 genes per Mb). To test whether this reduction was significant, we randomly reassigned pericentromeric regions in the human genome and assessed exon content at 2 and 5 Mb. A significant reduction was observed at 2 Mb ($P = 0.0008$) but not at 5 Mb ($P = 0.07$).

Although transcriptional activity within pericentromeric regions is uniformly reduced, the transposition of gene-rich euchromatic segments and the rapid evolutionary turnover of such regions creates the potential for the formation of new transcripts^{24,25}. We identified a total of 28 genes/messenger RNAs that had been completely duplicated within pericentromeric regions and for which there was evidence of transcription (as determined by best EST placement) (Supplementary Table 12 and Fig. 7). In addition to complete gene duplications, two other types of transcript innovations have been noted within pericentromeric regions: 'fusion' transcripts formed by the splicing of exons from two different

duplcon modules, and 'exapted' transcripts which acquire one or more exons outside of the ancestral duplicated region. We identified a total of 11 fusion and 17 exapted transcripts representing 28 novel transcript clusters. Eleven of these (Supplementary Table 12, Supplementary Fig. 7) were associated with predicted genes with open reading frames and may therefore represent emerging genes.

We selected 16 distinct pericentromeric genes, mRNA and/or ESTs where there was evidence of either exon fusion or exaptation for further expression analysis. We specifically designed polymerase chain reaction with reverse transcription (RT-PCR) assays at the site of fusion/exaptation and tested a panel of eight tissues (Supplementary Table 13). Almost all assays (15 out of 16) amplified complementary DNA from the testis and more than half showed evidence of transcription (9 out of 16) from the ovary. Interestingly, 7 of the assays were exclusively expressed from the testis. These data suggest that germline tissues are much more likely to express novel



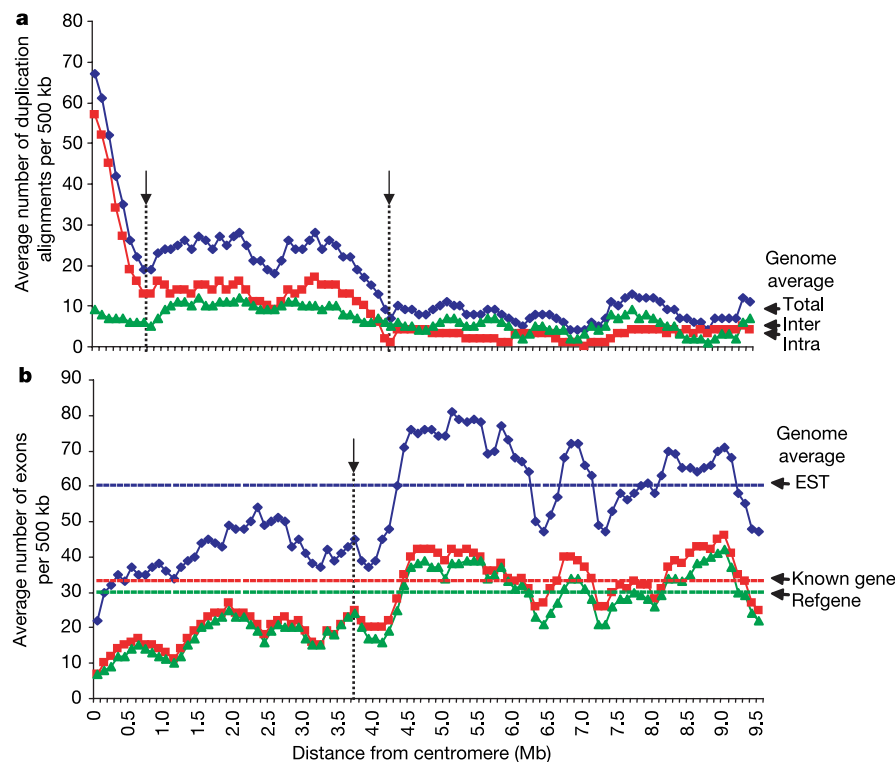


Figure 3 Sequence properties of centromeric transition regions. We computed a series of sequence properties (duplication content and exon density) in 500-kb windows (100-kb increments) for the first 10 Mb of each human chromosome arm beyond the centromere. The figures are based on the average for 30 pericentromeric regions where a large block of satellite sequence has been identified at the most proximal position. **a**, Plot of the average number of duplication alignments (blue diamonds, total) for both interchromosomal (red squares) and intrachromosomal (green triangles) duplications. Significant changes in interchromosomal duplication and overall duplication content are

noted at ~1 and 4.5 Mb, respectively. **b**, The average number of exons based on analysis of ESTs (blue diamonds), known genes (red squares) and RefSeq genes (green triangles) are shown. Each EST, known gene and RefSeq gene is placed uniquely based on highest sequence similarity scores. A significant decrease ($P = 0.008$) in exon density is observed at 2 Mb but not 5 Mb ($P = 0.07$). ESTs or genes with multiple tied placements are counted only once. See Supplementary Fig. 1 for other sequence properties and a breakdown by individual chromosome.

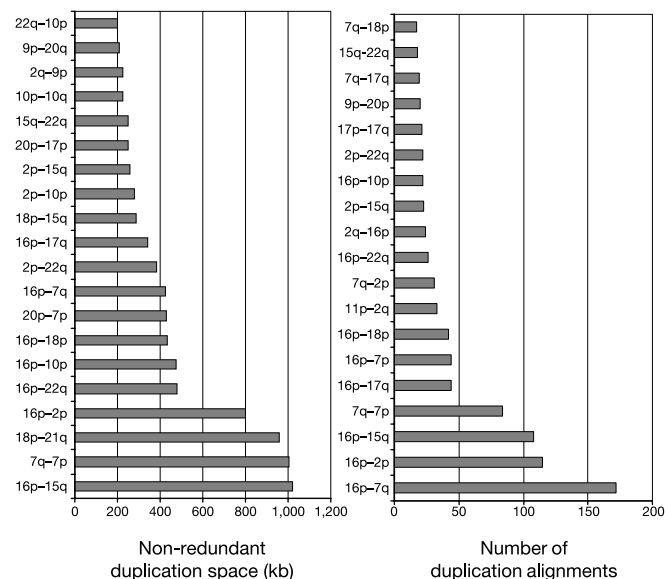


Figure 4 Cohorts of pericentromeric duplication. The histograms show the pericentromeric regions by cytogenetic location that are most preferentially associated by duplication. The top 20 regions are arranged based on the largest amount of shared duplicated sequence within 5 Mb of the centromere. Pericentromeric regions still represent work in progress. As additional sequence is generated, the rank order for specific pericentromeric regions may change.

pericentromeric transcripts than any other human tissue. Different packaging constraints of pericentromeric chromatin in germline tissue may contribute to this effect.

In summary, our analysis indicates that most human pericentromeric regions have been subjected to a complex series of duplications during the course of evolution (Fig. 2) with a gradient effect of interchromosomal duplications biased towards the centromere. Our delineation of ancestral donor sequences allows us to estimate a lower-bound rate for this process. We identified 271 ancestral segmental duplications to 43 pericentromeric regions over an estimated 40 million years of human evolution. We calculate an effective fixation rate of six–seven transposition events per million years. Subsequent pericentromeric duplications of these ancestral loci predict that the rate of duplication among non-homologous chromosomes was at least three times higher with an average fixation rate of about 20 events per million years. We believe that these two estimates are conservative as not all ancestral segments could be identified by mouse synteny analysis. If the total number of pericentromeric duplication alignments (8,343) is used as a surrogate for duplication/rearrangement events, the estimate may be as high as 206 duplication/rearrangement events per million years. Although the rate of pericentromeric duplication has been extensive, only a few juxtapositions of ancestral cassettes have created new transcripts. On the basis of our analysis, we estimate that a novel or mosaic transcript may have emerged through pericentromeric duplication once every million years of evolution. The fate and function of such evolutionary novelties remain to be determined. □

Table 1 Distribution of ancestral duplicons among pericentromeric regions

Chromosome arm	Ancestral duplicons				Pericentromeric regions (5 Mb)			
	Observed	Expected (± 2 s.d.)	Poisson (<i>P</i>)	Simulation	Observed	Expected (± 2 s.d.)	Poisson (<i>P</i>)	Simulation
Chr 1p†	9	11.7 $\pm$ 6.7	0.0939	2,587	3	18.7 $\pm$ 7.9	<u>0.0000</u>	0
Chr 1q	11	10.3 $\pm$ 6.3	0.1166	4,594	13	14.2 $\pm$ 7.1	0.1042	4,314
Chr 2p†	11	9.0 $\pm$ 5.9	0.0970	2,920	41	14.4 $\pm$ 7.0	<u>0.0000</u>	0
Chr 2q†	11	14.8 $\pm$ 7.4	0.0698	1,904	6	18.3 $\pm$ 7.9	<u>0.0006</u>	2
Chr 3p†	4	8.9 $\pm$ 5.8	0.0357	547	0	19.0 $\pm$ 8.0	<u>0.0000</u>	0
Chr 3q†	7	10.3 $\pm$ 6.3	0.0821	1,912	5	18.9 $\pm$ 8.0	<u>0.0001</u>	1
Chr 4p††	13	4.5 $\pm$ 4.1	<u>0.0006</u>	7	5	18.3 $\pm$ 8.0	<u>0.0002</u>	0
Chr 4q†	10	13.8 $\pm$ 7.3	0.0701	1,815	0	18.9 $\pm$ 8.0	<u>0.0000</u>	0
Chr 5p†	7	4.3 $\pm$ 4.1	0.0732	1,415	2	19.0 $\pm$ 8.1	<u>0.0000</u>	0
Chr 5q†	8	13.1 $\pm$ 7.0	0.0440	874	4	18.9 $\pm$ 8.0	<u>0.0000</u>	0
Chr 6p	2	5.6 $\pm$ 4.6	0.0580	779	9	18.6 $\pm$ 8.0	0.0061	79
Chr 6q†	9	10.7 $\pm$ 6.4	0.1142	3,609	5	18.7 $\pm$ 8.1	<u>0.0001</u>	2
Chr 7p†	4	5.4 $\pm$ 4.6	0.1600	3,666	41	18.7 $\pm$ 8.0	<u>0.0000</u>	0
Chr 7q†	10	9.5 $\pm$ 6.0	0.1235	4,799	61	18.4 $\pm$ 7.9	<u>0.0000</u>	0
Chr 8p†	1	4.0 $\pm$ 4.0	0.0733	925	0	19.0 $\pm$ 8.1	<u>0.0000</u>	0
Chr 8q*	0	9.7 $\pm$ 6.1	<u>0.0001</u>	1	7	18.5 $\pm$ 7.9	0.0013	7
Chr 9p†	3	4.0 $\pm$ 4.0	0.1954	4,212	31	14.2 $\pm$ 7.2	<u>0.0000</u>	0
Chr 9q†	2	7.0 $\pm$ 5.2	0.0223	272	37	15.1 $\pm$ 7.2	<u>0.0000</u>	0
Chr 10p	4	3.5 $\pm$ 3.7	0.1888	4,673	24	18.7 $\pm$ 8.1	0.0408	1,194
Chr 10q	15	9.1 $\pm$ 5.8	0.0208	352	20	17.8 $\pm$ 7.9	0.0776	3,185
Chr 11p	2	4.8 $\pm$ 4.3	0.0948	1,387	14	17.6 $\pm$ 7.8	0.0715	2,150
Chr 11q†	9	7.8 $\pm$ 5.5	0.1207	3,704	3	18.9 $\pm$ 8.2	<u>0.0000</u>	0
Chr 12p†	6	3.0 $\pm$ 3.4	0.0504	866	7	18.9 $\pm$ 8.0	<u>0.0011</u>	9
Chr 12q†	8	9.3 $\pm$ 6.0	0.1269	4,056	6	19.0 $\pm$ 8.0	<u>0.0004</u>	0
Chr 13q†	10	9.4 $\pm$ 6.0	0.1228	4,667	37	18.9 $\pm$ 8.0	<u>0.0001</u>	7
Chr 14q†	12	8.5 $\pm$ 5.7	0.0604	1,454	5	18.9 $\pm$ 8.1	<u>0.0001</u>	1
Chr 15q†	7	7.9 $\pm$ 5.5	0.1413	4,526	73	17.8 $\pm$ 7.9	<u>0.0000</u>	0
Chr 16p†	2	3.2 $\pm$ 3.6	0.2087	3,798	92	18.1 $\pm$ 7.9	<u>0.0000</u>	0
Chr 16q†	7	4.0 $\pm$ 3.9	0.0595	1,021	3	18.8 $\pm$ 8.0	<u>0.0000</u>	0
Chr 17p†	6	1.7 $\pm$ 2.6	0.0061	73	51	18.3 $\pm$ 7.9	<u>0.0000</u>	0
Chr 17q*	18	5.2 $\pm$ 4.5	<u>0.0000</u>	0	30	18.9 $\pm$ 8.1	0.0047	76
Chr 18p	1	1.0 $\pm$ 2.0	0.3679	6,498	20	18.9 $\pm$ 8.1	0.0862	4,344
Chr 18q†	3	5.6 $\pm$ 4.7	0.1082	1,907	0	19.0 $\pm$ 8.1	<u>0.0000</u>	0
Chr 19p†	6	2.0 $\pm$ 2.7	0.0120	3,183	6	18.9 $\pm$ 8.0	<u>0.0004</u>	0
Chr 19q†	3	2.7 $\pm$ 3.2	0.2205	5,092	0	18.9 $\pm$ 8.1	<u>0.0000</u>	0
Chr 20p	4	2.2 $\pm$ 2.9	0.1082	1,769	14	18.9 $\pm$ 8.1	0.0522	1,360
Chr 20q†	6	3.0 $\pm$ 3.4	0.0504	779	4	15.1 $\pm$ 7.2	<u>0.0006</u>	5
Chr 21q	3	2.9 $\pm$ 3.4	0.2237	5,605	23	19.0 $\pm$ 8.0	0.0556	1,890
Chr 22q	6	3.0 $\pm$ 3.5	0.0504	824	29	18.1 $\pm$ 8.0	0.0045	82
Chr Xp†	4	5.3 $\pm$ 4.6	0.1641	3,882	3	19.0 $\pm$ 8.0	<u>0.0000</u>	0
Chr Xq†	7	9.1 $\pm$ 5.9	0.1145	3,094	7	18.9 $\pm$ 8.0	<u>0.0011</u>	11

The number of ancestral duplicons originating within each chromosome arm (donor loci) and the number of observed duplicate copies observed within each 5-Mb pericentromeric region (acceptor loci) were counted (Methods). The chromosomal arms in which pericentromeric regions have a significant transition block (>10 kb) of alpha-satellite or other centromeric satellite DNA are in bold font. We observed 271 ancestral regions duplicated to 741 pericentromeric locations in the human genome. The expected number (± 2 s.d.) indicates the range for expected number of ancestral duplications or pericentromeric duplications by random distribution model. We tested for significant departures from a random genome distribution model by both Poisson sampling and by simulation. Significance values were corrected for multiple tests (Bon-Ferroni $P < 0.0012$). Regions significantly enriched in duplication are italicized, whereas regions 'protected' for duplication are underlined. The 95% confidence interval indicates the range of the expected number of ancestral or pericentromeric duplications based on a simulation of 10,000 replicates. The simulation reports the number of tests (from 10,000 replicates) that were equal to or exceeded the observed count for regions that were enriched, or the number of tests that were equal to or less than the observed count for regions that were protected. Most ancestral donor loci are randomly distributed but pericentromeric regions show a highly nonrandom distribution due largely to secondary duplications among specific cohorts of chromosomes.

*Significant departure from a random distribution model for donors.

†Significant departure from a random distribution model for acceptors.

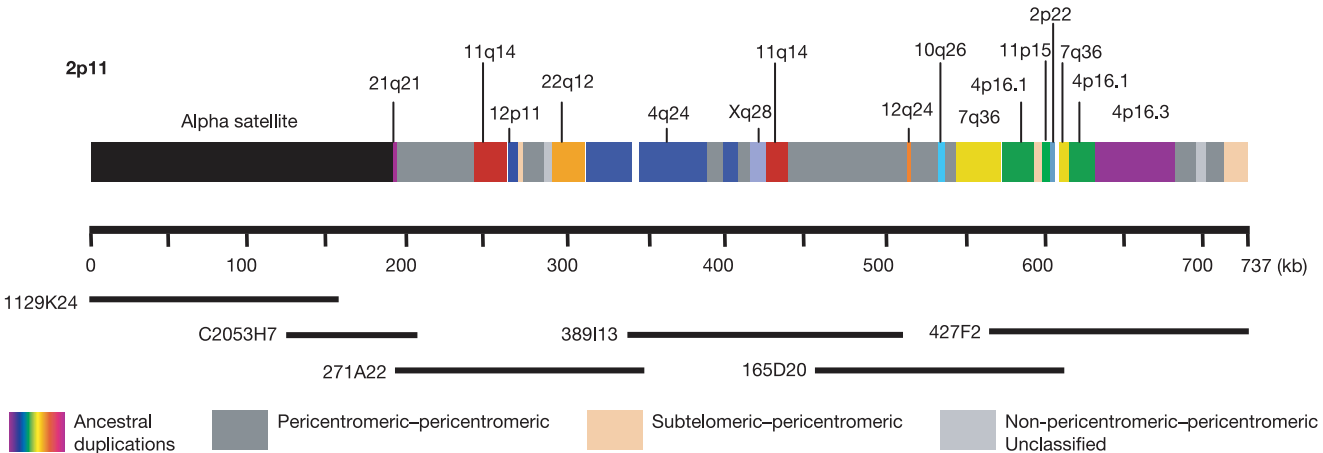


Figure 5 Ancestral duplicons within 2p11. The modular organization of a pericentromeric region, 2p11, is depicted based on the classification of the underlying pairwise alignments. Duplicated segments that originate outside of the pericentromeric region, termed ancestral duplicons, are shown in colour (ancestral cytogenetic band locations are delineated). Unshaded regions correspond to regions where no underlying duplication could be detected. The minimal tiling path of large-insert BAC clones is drawn to scale

below each line. A 737-kb validated sequence contig that provides the first autosomal transition into higher-order alpha-satellite repeat DNA is shown here. Approximately 98% of this region is composed of duplicated material of which 57% can be traced back to non-pericentromeric regions of the genome. These correspond to 13 ancestral duplicons of which 9 were experimentally confirmed by non-human primate analyses.

Methods

Characterization and validation of pericentromeric DNA

The definition of pericentromeric DNA and its boundary is arbitrary. Noticeable changes in exon density, satellite repeat and interchromosomal duplication content were observed at 2.6 and 5.2 Mb (Fig. 3). In this study, therefore, we limited our analysis of pericentromeric DNA to both 2 and 5 Mb. To facilitate the assembly and to assess the quality of these regions, we performed STS content analysis, FISH and detailed characterization of alpha-satellite repeat content (Supplementary Methods). The origin of ancestral duplications for the 2p11 region was determined based on detailed comparative and phylogenetic analysis between humans and non-human primate species (Supplementary Methods). All genome-wide analyses were performed based on the July 2003 finished genome assembly (build34) and are available at <http://humanparalogy.gene.cwru.edu>.

Duplication analyses

We used a BLAST-based detection scheme⁶ to initially identify all pairwise similarities representing duplicated regions (≥ 1 kb and $\geq 90\%$ identity) within the finished sequence of the human genome (July 2003). A total of 25,239 welded pairwise alignments were generated, of which 8,343 mapped within 5 Mb of human pericentromeric DNA. To eliminate potential artefactual duplications due to misassembly, we considered only those alignments that could be confirmed using a second assembly-free method of detection (whole-genome shotgun sequence detection²⁴). To test the significance of the observed enrichment of duplicated sequences within pericentromeres (5 Mb around the centromere), a conservative model for the duplication process by randomly reassigning contiguous blocks of duplicated sequence to new locations was used. In order to detect more divergent duplications, a second all-by-all genome BLASTZ discontinuous search was performed within the finished genome to recover more divergent ($>75\%$) and shorter (>250 bp) alignments (W.M., unpublished data) (Supplementary Methods).

Ancestral origin of pericentromeric duplications

Using mouse synteny data, we could classify 741 of the 8,343 pericentromeric duplication alignments as ancestral and thereby define the directionality of the duplication event (see Supplementary Methods for details). Regions were clustered if duplcon subgroups mapped <100 kb of one another and the sequence divergence of pericentromeric alignments showed $<2\%$ difference. These ancestral duplications correspond to 8.1 Mb or 30% of all pericentromeric duplicated base pairs (8.1 out of 26 Mb). The remaining 18 Mb corresponded to pericentromeric duplications that did not have a euchromatic origin and/or regions where insufficient mouse-human synteny data existed to claim directionality of the duplication event. Ancestral loci were examined for the presence of intron-exon structure based on known genes (<http://genome.ucsc.edu>). A gene segment was considered if at least a single exon and intron could be identified within the duplicated segment. The Y chromosome is not included in this analysis due to the lack of mouse synteny information.

Gene and transcript analysis

We examined transcriptional potential by considering both the number of genes and exons in 500-kb windows (including gaps) sliding by 100-kb increments from the centromere for each chromosome. Four sets of data were considered: Refseq annotated genes (28,452), known annotated genes (38,482), mRNAs (130,762) and human ESTs with intron/exon structure (2.3 million). Both best-placement and tied genes/transcripts were distinguished based on BLAT score criteria (<http://www.genome.ucsc.edu>). Exon density was defined as the number of non-overlapping exons identified within a genomic window. We separately examined mRNAs and genes that mapped to duplicated regions near the centromere (5 Mb), considering only those where the sequence identity between genome and cDNA exceeded 99%. A total of 25 known and Refseq genes and 31 mRNA clusters, with intron/exon structure, were identified within the duplicated regions. The calculation for novel transcripts or genes was estimated based on sequence divergence of genomic sequence. Marmoset DNA shows $\sim 11\text{--}12\%$ (ref. 34) divergence from human and they are estimated to have diverged 35–40 million years ago. On the basis of sequence divergence, the segmental duplications observed here probably do not exist in marmoset and therefore the 28 novel genes/mRNAs arose in the last 35 million years of evolution (28 genes per 35 million years = ~ 1 gene per million years).

Received 21 February; accepted 2 July 2004; doi:10.1038/nature02806.

- Nagaki, K. *et al.* Sequencing of a rice centromere uncovers active genes. *Nature Genet.* **36**, 138–145 (2004).
- Horvath, J. *et al.* Molecular structure and evolution of an alpha/non-alpha satellite junction at 16p11. *Hum. Mol. Genet.* **9**, 113–123 (2000).
- Jackson, M. Duplicate, decouple, disperse: the evolutionary transience of human centromeric regions. *Curr. Opin. Genet. Dev.* **13**, 629–635 (2003).
- International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Eichler, E. E. Masquerading repeats: Paralogous pitfalls of the Human Genome. *Genome Res.* **8**, 758–762 (1998).
- Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
- Istrail, S. *et al.* Whole-genome shotgun assembly and comparison of human genome assemblies. *Proc. Natl Acad. Sci. USA* **101**, 1916–1921 (2004).
- Bailey, J. A. *et al.* Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* **70**, 83–100 (2002).
- Bailey, J. A., Giu, L. & Eichler, E. E. An Alu transposition model for the origin and expansion of

- human segmental duplications. *Am. J. Hum. Genet.* **73**, 823–834 (2003).
- Horvath, J. E., Bailey, J. A., Locke, D. P. & Eichler, E. E. Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum. Mol. Genet.* **10**, 2215–2223 (2001).
- Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K. & Willard, H. F. Genomic and genetic definition of a functional human centromere. *Science* **294**, 109–115 (2001).
- Horvath, J. E. *et al.* Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol. Biol. Evol.* **20**, 1463–1479 (2003).
- Ventura, M. *et al.* Neocentromeres in 15q24–26 map to duplcons which flanked an ancestral centromere in 15q25. *Genome Res.* **13**, 2059–2068 (2003).
- Yunis, J. J. & Prakash, O. The origin of man: a chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982).
- Baldini, A. *et al.* An aphoid DNA sequence conserved in all human and great ape chromosomes: evidence for ancient centromeric sequences at human chromosomal regions 2q21 and 9q13. *Hum. Genet.* **90**, 577–583 (1993).
- Liu, G. *et al.* Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res.* **13**, 358–368 (2003).
- Ventura, M., Archidiacono, N. & Rocchi, M. Centromere emergence in evolution. *Genome Res.* **11**, 595–599 (2001).
- Eder, V. *et al.* Chromosome 6 phylogeny in primates and centromere repositioning. *Mol. Biol. Evol.* **20**, 1506–1512 (2003).
- Eichler, E. E. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.* **17**, 661–669 (2001).
- Locke, D. P. *et al.* Molecular evolution of the human chromosome 15 pericentromeric region. *Cytogenet. Genome Res.* (in the press).
- Golfier, G. *et al.* The 200-kb segmental duplication on human chromosome 21 originates from a pericentromeric dissemination involving human chromosomes 2, 18 and 13. *Gene* **312**, 51–59 (2003).
- Courseaux, A. *et al.* Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation. *Genome Res.* **13**, 369–381 (2003).
- Horvath, J., Schwartz, S. & Eichler, E. The mosaic structure of a 2p11 pericentromeric segment: A strategy for characterizing complex regions of the human genome. *Genome Res.* **10**, 839–852 (2000).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- Guy, J. *et al.* Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10p. *Genome Res.* **13**, 159–172 (2003).
- Hillier, L. W. *et al.* The DNA sequence of human chromosome 7. *Nature* **424**, 157–164 (2003).
- Crosier, M. *et al.* Human paralogs of KIAA0187 were created through independent pericentromeric-directed and chromosome-specific duplication mechanisms. *Genome Res.* **12**, 67–80 (2002).
- Jackson, M. S. *et al.* Sequences flanking the centromere of human chromosome 10 are a complex patchwork of arm-specific sequences, stable duplications, and unstable sequences with homologies to telomeric and other centromeric locations. *Hum. Mol. Genet.* **8**, 205–215 (1999).
- Locke, D. P. *et al.* Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* **13**, 347–357 (2003).
- Thomas, J. W. *et al.* Pericentromeric duplications in the laboratory mouse. *Genome Res.* **13**, 55–63 (2003).
- Tuzun, E., Bailey, J. & Eichler, E. E. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* **14**, 493–506 (2004).
- Bailey, J. A., Baertsch, R., Kent, W. J., Haussler, D. & Eichler, E. E. Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**, R23 (2004).
- Copenhaver, G. P. *et al.* Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**, 2468–2474 (1999).
- Schneider, H. *et al.* Molecular phylogeny of the New World monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences. *Am. J. Phys. Anthropol.* **100**, 153–179 (1996).
- Willard, H. F. & Wayne, J. S. Chromosome-specific subsets of human alpha satellite DNA: analysis of sequence divergence within and between chromosomal subsets and evidence for an ancestral pentameric repeat. *J. Mol. Evol.* **25**, 207–214 (1987).
- Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V. & Yurov, Y. Alpha-satellite DNA of primates: old and new families. *Chromosoma* **110**, 253–266 (2001).
- Lee, C., Wevrick, R., Fisher, R. B., Ferguson-Smith, M. A. & Lin, C. C. Human centromeric DNAs. *Hum. Genet.* **100**, 291–304 (1997).
- An International System for Human Cytogenetic Nomenclature, High resolution-banding. *Cytogenet. Cell Genet.* **31**, 1–23 (1981).
- Leem, S. H. *et al.* Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. *Genome Res.* **14**, 239–246 (2004).
- Rudd, M. K. & Willard, H. F. Analysis of the centromeric regions of the human genome assembly. *Trends Genet.* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the large-scale sequencing centres (Baylor College of Medicine, Cold Spring Harbor Laboratory, Genome Therapeutics Corporation, Harvard Partners Genome Center, Joint Genome Institute, The NIH Intramural Sequencing Center, The UK-MRC Sequencing Consortium, The University of Oklahoma Advanced Center for Genome Technology, The University of Texas Southwest, The Whitehead Institute for Biomedical Research, The Washington University Genome Sequencing Center and the Wellcome Trust Sanger Institute) for access to all large-scale finished sequence, genome assembly and trace sequence data from the human genome before publication. This work was supported by grants from NIH and DOE to E.E.E. and grants from P.R.I.N.C.E., MURST and Telethon to M.R.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.E.E. (eee@gs.washington.edu).

Discovery of five irregular moons of Neptune

Matthew J. Holman¹, J. J. Kavelaars², Tommy Grav^{1,3}, Brett J. Gladman⁴, Wesley C. Fraser⁵, Dan Milisavljevic⁵, Philip D. Nicholson⁶, Joseph A. Burns⁶, Valerio Carruba⁶, Jean-Marc Petit⁷, Philippe Rousselot⁷, Oliver Mousis⁷, Brian G. Marsden¹ & Robert A. Jacobson⁸

¹Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

²National Research Council of Canada, 5071 West Saanich Road, Victoria, British Columbia V9E 2E7, Canada

³University of Oslo, Institute of Theoretical Astrophysics, Postbox 1029 Blindern, 0315 Oslo, Norway

⁴Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada

⁵Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada

⁶Department of Astronomy, Cornell University, Ithaca, New York 14853, USA

⁷Observatoire de Besançon, BP 1615, 25010 Besançon Cedex, France

⁸Jet Propulsion Laboratory, MS 301-150, 4800 Oak Grove Drive, Pasadena, California 91109, USA

Each giant planet of the Solar System has two main types of moons. 'Regular' moons are typically larger satellites with prograde, nearly circular orbits in the equatorial plane of their host planets at distances of several to tens of planetary radii. The 'irregular' satellites (which are typically smaller) have larger orbits with significant eccentricities and inclinations. Despite these common features, Neptune's irregular satellite system, hitherto thought to consist of Triton and Nereid, has appeared unusual. Triton is as large as Pluto and is postulated to have been captured from heliocentric orbit; it traces a circular but retrograde orbit at 14 planetary radii from Neptune. Nereid, which exhibits one of the largest satellite eccentricities, is believed to have been scattered from a regular satellite orbit to its present orbit during Triton's capture^{1,2}. Here we report the discovery of five irregular moons of Neptune, two with prograde and three with retrograde orbits. These exceedingly faint (apparent red magnitude $m_R = 24.2\text{--}25.4$) moons, with diameters of 30 to 50 km, were presumably captured by Neptune.

Recent searches for neptunian moons have employed digital techniques^{3–6}, but have revealed no new neptunian moons. The lack of discoveries has been interpreted⁶ as supporting the theory of a violent destruction of the neptunian outer satellite system when Triton was captured and tidally circularized^{1,2}. However, recent discoveries of small jovian⁷, saturnian⁸ and uranian⁹ irregular satellites suggested that previously undetected neptunian satellites might be found just beyond the earlier surveys' detection thresholds⁸.

Thus, in 2001, we searched for fainter neptunian moons, using a more sophisticated method. With the Cerro Tololo Inter-American Observatory 4-m and Canada-France-Hawaii 3.6-m telescopes, we searched 1.4 square degrees centred on Neptune (Fig. 1). The volume near a planet where satellites are dynamically stable is roughly given by the 'Hill sphere', within which the planet's gravity overcomes solar tidal perturbations¹⁰. The radius of the Hill sphere is $R_H = a_p(\mu/3)^{1/3}$, where a_p is the planet's semimajor axis and μ is the ratio of the planet's mass to the Sun's. For Neptune, $R_H = 1.15 \times 10^{11}$ km (0.77 AU) or 1.5° viewed from Earth. Detailed 10-Myr numerical integrations show that prograde satellites (those orbiting in the same sense as the planet orbits the Sun) of Neptune are stable to distances of $\sim 0.4 R_H$, and retrograde satellites (those orbiting in the opposite sense) are stable to $\sim 0.7 R_H$, depending upon the satellite's eccentricity and inclination¹¹. Thus, our search

was sensitive to nearly all prograde orbits. However, some retrograde satellites might lie beyond our search region (Fig. 1). We repeated our search in August 2002 and August 2003, with CTIO's Blanco telescope, in order to recover our satellites.

On each search night, we took 20–25 eight-minute exposures of one or more of our four search regions. We imaged through a "VR" filter, which is centred between the V and R bands, and is approximately 100 nm wide¹². We also acquired images of photometric standard fields¹³ to transform the VR observations to the Kron R photometric system.

For our searches we adapted a pencil-beam technique developed to detect faint Kuiper belt objects^{12,14}. The detection of faint, moving objects in a single exposure is limited by the object's motion. Long exposures spread the signal from the object in a trail; the atmospheric conditions and the quality of the telescope optics restrict the useful exposure time to the period in which the object traverses the width of the point-spread function. However, as the apparent rate and direction at which Neptune moves across the sky are known, we shift the successive images in software to compensate for that motion. We then combine the exposures to produce a single deep image. The signal from objects moving near Neptune's rate and direction collects into a point-like image. We repeat this procedure for the range of rates and directions of apparent motion that is consistent with bound satellites within our search fields. Subtraction of a template image, scaled to match the varying background and point-spread function, from each individual exposure, before it

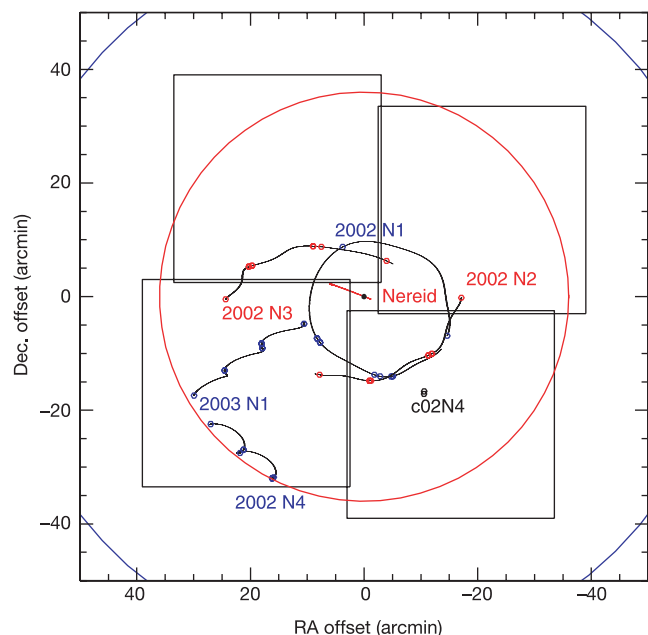


Figure 1 Search regions. The rosette of black squares indicate the $36' \times 36'$ CTIO Mosaic-II camera fields surrounding Neptune (at centre) searched in August 2002 and 2003. Our search in 2001 had a similar field pattern, but the northwest field was searched with the CFH12K camera ($28' \times 42'$). This layout covers a roughly circular region, while minimizing field overlap and avoiding significant scattered light contamination from Neptune. The open symbols mark the observed satellite positions with respect to Neptune. The traces of offset positions of the irregular satellites announced here, as well as that of Nereid, are also shown. Owing to the changing perspective from the Earth, these do not appear as segments of ellipses. The red (blue) solid circle roughly indicates the stability limit of prograde (retrograde) ecliptic satellites. The prograde and retrograde satellites are also labelled in red and blue, respectively. The labels are placed near the offset positions at the time of discovery (July or August 2001 for all but S/2002 N 4 and c02N4). In summer 2001, S/2002 N 4 was outside our search region and c02N4 was not detected. The observed positions of the unrecovered satellite candidate, c02N4, are indicated, but it is not classified as prograde or retrograde, as its orbit is undetermined.

Table 1 The time-averaged orbital elements of Neptune's irregular satellites

Satellite	<i>R</i> (mag)	Diam. (km)	<i>a</i> _{ave} (10 ⁶ km)	<i>e</i> _{ave}	<i>e</i> _{min}	<i>e</i> _{max}	<i>i</i> _{ave} (deg)	<i>i</i> _{min}	<i>i</i> _{max}	<i>P</i> (yr)	<i>P</i> _{coll}
Triton	13.5	2707	0.35	0.00			156.8			0.016	
Nereid	19.9	340	5.53	0.75	0.73	0.76	9.8	3.9	15.1	0.986	
S/2002 N 1	24.2	54	16.6	0.43	0.11	0.93	114.9	106.0	144.2	5.14	0.410
S/2002 N 2	25.4	31	22.3	0.27	0.06	0.63	50.4	39.2	56.3	7.98	0.015
S/2002 N 3	25.0	37	23.5	0.36	0.25	0.49	35.9	29.6	41.9	8.67	—
S/2002 N 4	24.7	43	48.6	0.39	0.11	0.76	137.4	126.6	151.0	25.77	—
S/2003 N 1	25.1	36	47.6	0.49	0.13	0.93	125.1	112.4	148.7	26.58	0.013
c02N4	25.3	33	25.1								

The apparent *R*-band magnitudes and diameters, along with orbital elements, current osculating orbital period (*P*), and probability of collision with Nereid (over 4.5 Gyr; *P*_{coll}) are given. The mean, minimum and maximum orbital elements, with respect to the J2000 ecliptic and equinox, are based on our numerical integrations. The diameter estimates of the newly discovered satellites assume a 6% geometric albedo. For c02N4, the projected distance between it and Neptune at the time of discovery is listed. The semimajor axes (*a*) of the newly discovered satellites are all significantly larger than that of Nereid but show little variation with time. However, the eccentricities (*e*) and inclination (*i*) undergo large variations. At their maximum the eccentricities of S/2002 N 1 and S/2003 N 1 exceed that of Nereid.

is shifted and combined¹⁵, eliminates stars and other stationary sources of light that would otherwise appear as trails in these shifted and combined images. We apply two different detection algorithms to the shifted and combined images to identify flux sources¹⁶. These two algorithms have different false-detection characteristics; retaining only those sources detected by both algorithms eliminates a significant fraction of the spurious detections associated with cosmic-ray contamination and background fluctuations. Before applying our search algorithms to the data, we implant a large number of artificial objects, with various magnitudes, moving at different directions and rates, to calibrate our search. Our detection efficiency is found to be independent of the object's motion. Recovery observations employed both standard three image and pencil-beam techniques. Because our 12–15 August 2002 search had the most uniform conditions, we characterize our survey with it. Fitting an empirical detection efficiency function, $\epsilon = \frac{1}{2}\epsilon_{\max}\{1 - \tanh[(m_R - m_{50})/w]\}$, we find a 50% detection threshold of $m_{50} = 25.5$, a detection rate of $\epsilon_{\max} = 97\%$ for objects much brighter than the detection limit, and a transition from high to low detection efficiency over roughly $2w = 0.6$ mag.

Three searches yielded five new neptunian irregulars, four of which were identified in our 2001 data. All five were detected by our algorithms in the August 2002 images; however, one (S/2003 N 1) was initially overlooked during our visual inspections and was first reported by others¹⁷. Additional follow-up observations were conducted with the VLT, Magellan-II, Palomar 5-m, and Nordic Optical telescopes. These five satellites have been announced, based on recoveries in 2003, and have been re-observed in June 2004. A sixth candidate, which we designated 'c02N4', was discovered on 14 August 2002 and seen again at the VLT on 3 September 2002. Further attempts to recover this object failed. Although c02N4 is possibly a Centaur, it moved very little relative to Neptune between August and September, more consistent with it being a satellite. We also note that S/2002 N 1 was recently identified, based on our orbital solutions, in images from a 1999 search of the full Hill sphere of Neptune to a limit of $m_R = 24.3$ (ref. 6). Of the newly discovered neptunians, its orbit is now the best determined, with observations covering nearly a full orbital period. The other new neptunian satellites are much fainter than the detection threshold of that search.

Irregular satellites are widely believed to be captured from heliocentric orbits. Objects on planet-crossing orbits, with low speeds relative to the planet, can be temporarily captured into planetocentric orbit. However, apart from striking the planet, such temporarily captured bodies typically return to a heliocentric orbit in 10–100 orbits unless enough orbital energy is dissipated to make the capture permanent^{18,19}. Several dissipation mechanisms have been proposed: (1) a sudden increase in the planetary mass through accretion of nearby material¹⁸; (2) collision or gravitational interaction with an extant moon or with another temporarily captured object²⁰; (3) gas drag in an extended envelope²¹ or disk¹⁹ surrounding the still-forming planet; and (4) dynamical friction from a

background of small outer Solar System bodies^{22,23}.

Although the new satellites' orbits will undoubtedly be revised as more observations become available, we have numerically integrated their preliminary orbits for 10 Myr, including the Sun and giant planets as perturbers, to investigate their long-term dynamical stability²⁴. Table 1 lists the best-fit mean orbital elements of the satellites. We find that S/2002 N 2 is in the Kozai resonance^{11,25,26}. All of the new neptunians are stable on the timescale of the integrations. Figure 2 shows the eccentricity, inclination and semimajor axis for the neptunian irregular satellites with determined orbits, along with the results of numerical integrations to determine dynamical stability.

Neptune's irregulars are probably the remnants of a system collisionally and gravitationally altered by Triton and Nereid, as well as by the solar tidal potential. Curiously, all five new neptunian irregulars with well-determined orbits have minimum pericentre distances near the apocentre of Nereid; S/2002 N 1 crosses the orbit of Nereid most deeply at its minimum pericentre. This is consistent with an isotropic distribution of initial orbits, subsequently sculpted by gravitational perturbations. Collisions between Nereid and nearby satellites are probable¹¹. We estimate the probability of collision between the new neptunian irregulars and Nereid, assuming fixed orbital semimajor axes, eccentricities and inclinations with

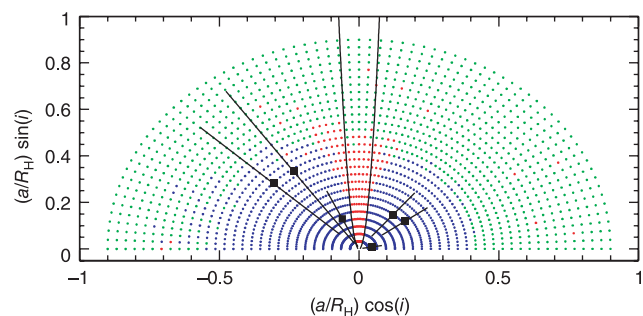


Figure 2 Dynamical stability of neptunian moons. The points indicate the initial semimajor axes and inclinations of test particles in orbit about Neptune, with the colours indicating the outcome of their numerical integration. Blue points denote particles that survived for the full 10-Myr integration. Red and green points denote the particles that moved inside the orbit of Triton ($r \approx 0.0024$ AU) or outside 1.5 times the Hill sphere of Neptune ($r \approx 1.167$ AU), respectively. A symplectic *n*-body map²⁴, modified for satellite orbits, was used. The gravitational perturbations of the Sun and giant planets were included, while those from Triton and Nereid were neglected. The test particles were started with semimajor axes of $a = 0.025$ – 0.70 AU ($\Delta a = 0.025$ AU), inclinations $i = 0$ – 180° ($\Delta i = 2.5^\circ$), eccentricity $e = 0.5$, and argument of pericentre $\omega = 90^\circ$. The longitudes of ascending node Ω and mean anomalies M were chosen randomly. Also shown are the mean orbital elements of the new neptunian irregular satellites with secure orbits, along with that of Nereid. The lines indicate the mean pericentric and apocentric distances of these moons from Neptune. The region within which the Kozai mechanism removes nearly polar orbits is indicated by the solid, nearly parabolic line^{11,25,26}.

random longitudes of ascending node and arguments of periape and mean anomalies²⁷. Variations in eccentricity strongly affect the collision probability for marginally crossing orbits. We account for this by averaging the collision probability over several eccentricity oscillations induced by the solar tide. The collision probability between S/2002 N 1 and Nereid over 4.5 Gyr is 0.41. Between Nereid and S/2002 N 2 and S/2003 N 1 these are 0.016 and 0.013, respectively. For S/2002 N 3 and S/2002 N 4, the collision probabilities with Nereid are negligible. Furthermore, the probability of collisions among the new neptunians is negligible. Nereid's Hill sphere, relative to Neptune, is roughly the size of Neptune itself. The likelihood of gravitational scattering is a factor of 2×10^4 larger than that of physical collision. The high relative velocity between Nereid and the new neptunian irregulars during the encounters eliminates the possibility of direct ejection. However, it is possible that repeated gentle gravitational scattering by Nereid has shaped the orbital distribution of neptunian irregular satellites.

The jovian and saturnian irregular satellites cluster in families with similar inclination and semimajor axis^{7,8}. Photometry of the jovian and saturnian irregular satellites shows that most satellite clusters have homogeneous colours, thus intimating that irregulars are collisional fragments of larger progenitors^{28,29}. The similarities between the orbits of S/2002 N 2 and S/2002 N 3, and between those of S/2003 N 1 and S/2002 N 4, suggest that such families might exist among the neptunian irregulars as well. The remaining retrograde satellite, S/2002 N 1, has an inclination similar to that of the other retrogrades, but its semimajor axis is significantly smaller. This has been identified as a characteristic of 'chaos-assisted capture'²². Alternatively, S/2002 N 1 might be a collisional fragment of Nereid, consistent with its large probability of collision with Nereid. Photometric observations of the S/2002 N 1 show that its optical colours are similar to Nereid's³⁰. Additional photometric observations of the new neptunian satellites to compare their colours, along with further searches for additional satellites, are needed to determine whether Neptune also hosts collisional families of irregular satellites. □

Received 2 February; accepted 12 July 2004; doi:10.1038/nature02832.

- McKinnon, W. B. On the origin of Triton and Pluto. *Nature* **311**, 355–358 (1984).
- Goldreich, P., Murray, N., Longaretti, P.-Y. & Banfield, D. Neptune's story. *Science* **245**, 500–504 (1989).
- Hogg, D. W. *et al.* A photographic search for satellites of Neptune. *Icarus* **107**, 304–310 (1994).
- Gladman, B. J. *et al.* Discovery of two distant irregular moons of Uranus. *Nature* **392**, 897–899 (1998).
- Brown, M. J. I. & Webster, R. L. A search for distant satellites of Neptune. *Publ. Astron. Soc. Austr.* **15**, 325–327 (1998).
- Gladman, B. *et al.* NOTE: The discovery of Uranus XIX, XX, and XXI. *Icarus* **147**, 320–324 (2000).
- Sheppard, S. S. & Jewitt, D. C. An abundant population of small irregular satellites around Jupiter. *Nature* **423**, 261–263 (2003).
- Gladman, B. *et al.* Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* **412**, 163–166 (2001).
- Kavelaars, J. J. *et al.* The discovery of faint irregular satellites of Uranus. *Icarus* **169**, 474–481 (2004).
- Hénon, M. Numerical exploration of the restricted problem. VI. Hill's case: non-periodic orbits. *Astron. Astrophys.* **9**, 24–36 (1970).
- Nesvorný, D., Alvarro, J. L. A., Dones, L. & Levison, H. F. Orbital and collisional evolution of the irregular satellites. *Astron. J.* **126**, 398–429 (2003).
- Allen, R. L., Bernstein, G. M. & Malhotra, R. The edge of the solar system. *Astrophys. J.* **549**, L241–L244 (2001).
- Landolt, A. U. UBVR photometric standard stars in the magnitude range 11.5–16.0 around the celestial equator. *Astron. J.* **104**, 340–371 (1992).
- Gladman, B., Kavelaars, J. J., Nicholson, P. D., Lored, T. J. & Burns, J. A. Pencil-beam surveys for faint trans-neptunian objects. *Astron. J.* **116**, 2042–2054 (1998).
- Alard, C. Image subtraction using a space-varying kernel. *Astron. Astrophys.* **144** (Suppl.), 363–370 (2000).
- Petit, J.-M., Holman, M., Scholl, H., Kavelaars, J. & Gladman, B. An automated moving object detection package. *Mon. Not. R. Astron. Soc.* **347**, 471–480 (2004).
- Sheppard, S. S. *et al.* Satellites of Neptune. *IAU Circ. No.* 8193 (2003).
- Heppenheimer, T. A. & Porco, C. New contributions to the problem of capture. *Icarus* **30**, 385–401 (1977).
- Čuk, M. & Burns, J. A. Gas-drag-assisted capture of Himalia's family. *Icarus* **167**, 369–381 (2004).
- Colombo, G. & Franklin, E. A. On the formation of the outer satellite groups of Jupiter. *Icarus* **15**, 186–189 (1971).
- Pollack, J. B., Burns, J. A. & Tauber, M. E. Gas drag in primordial circumplanetary envelopes—A mechanism for satellite capture. *Icarus* **37**, 587–611 (1979).
- Astakhov, S. A., Burbanks, A. D., Wiggins, S. & Farrelly, D. Chaos-assisted capture of irregular moons. *Nature* **423**, 264–267 (2003).

- Goldreich, P., Lithwick, Y. & Sari, R. Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* **420**, 643–646 (2002).
- Wisdom, J. & Holman, M. Symplectic maps for the n-body problem. *Astron. J.* **102**, 1528–1538 (1991).
- Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67**, 591–598 (1962).
- Carruba, V., Burns, J. A., Nicholson, P. D. & Gladman, B. J. On the inclination distribution of the jovian irregular satellites. *Icarus* **158**, 434–449 (2002).
- Kessler, D. J. Derivation of the collision probability between orbiting objects: The lifetimes of Jupiter's outer moons. *Icarus* **48**, 39–48 (1981).
- Grav, T., Holman, M. J., Gladman, B. J. & Aksnes, K. Photometric survey of the irregular satellites. *Icarus* **166**, 33–45 (2003).
- Rettig, T. W., Walsh, K. & Consolmagno, G. Implied evolutionary differences of the jovian irregular satellites from a BVR color survey. *Icarus* **154**, 313–320 (2001).
- Grav, T., Holman, M. J. & Fraser, W. Photometry of irregular satellites of Uranus and Neptune. Preprint astro-ph/0405605 at (<http://arXiv.org>) (2004).

Acknowledgements We thank M. Lecar for discussions, and D. Trilling for observing assistance at Magellan. T. Abbott (CTIO) volunteered to observe during Director's Discretionary time. CTIO is operated by the Association of Universities for Research in Astronomy, Inc. (AURA), under a cooperative agreement with the National Science Foundation as part of the National Optical Astronomy Observatories. The CFHT is operated by the National Research Council of Canada, the Centre National de la Recherche Scientifique de France and the University of Hawaii. The VLT is operated by the European Southern Observatory. This work was supported by NASA and the Smithsonian Institution.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.H. (mholman@cfa.harvard.edu).

Addition of nanoparticle dispersions to enhance flux pinning of the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ superconductor

T. Haugan¹, P. N. Barnes¹, R. Wheeler¹, F. Meisenkothen¹ & M. Sumpson²

¹Air Force Research Laboratory, Wright-Patterson AFB, Ohio 45433-7919, USA

²Department of Materials Science and Engineering, The Ohio State University, Columbus, Ohio 43210, USA

Following the discovery of type-II high-temperature superconductors in 1986 (refs 1, 2), work has proceeded to develop these materials for power applications. One of the problems, however, has been that magnetic flux is not completely expelled, but rather is contained within magnetic fluxons, whose motion prevents larger supercurrents. It is known that the critical current of these materials can be enhanced by incorporating a high density of extended defects to act as pinning centres for the fluxons^{3,4}. $\text{YBa}_2\text{Cu}_3\text{O}_7$ (YBCO or 123) is the most promising material for such applications at higher temperatures (liquid nitrogen)^{3–13}. Pinning is optimized when the size of the defects approaches the superconducting coherence length ($\sim 2\text{--}4\text{ nm}$ for YBCO at temperatures $\leq 77\text{ K}$) and when the areal number density of defects is of the order of $(H/2) \times 10^{11}\text{ cm}^{-2}$, where H is the applied magnetic field in tesla^{3,4}. Such a high density has been difficult to achieve by material-processing methods that maintain a nanosize defect, except through irradiation⁵. Here we report a method for achieving a dispersion of $\sim 8\text{-nm}$ -sized nanoparticles in YBCO with a high number density, which increases the critical current (at 77 K) by a factor of two to three for high magnetic fields.

The flux pinning enhancement of type II superconductors with defects have been studied in both copper-oxide high- and low-temperature superconductor materials^{3,4,14}. An areal number density of second-phase defects of over 10^{11} cm^{-2} was previously

achieved in NbTi type II low-temperature superconductors to increase the critical current, J_c (ref. 14). Dislocations are one class of defects in YBCO (hereafter referred to as 123) reportedly achieved in thin films with densities varying from 10^9 – 10^{11} cm $^{-2}$, depending on the deposition parameters^{15,16}. Dislocations were shown to be directly responsible for pinning in thin films¹⁵, but their density is still less than what is desired for complete flux pinning in the regimes of 2–6 T (refs 15, 16). To our knowledge, a density of second-phase defects exceeding 10^{10} cm $^{-2}$ has not yet been achieved for YBCO. In addition to high defect density, it is desirable to minimize the defect size, which maximizes the superconducting volume and allows a higher areal defect density to be reached.

In this work, second-phase YBa $_2$ CuO $_5$ (hereafter referred to as 211) particles of nanometre size were introduced by growth of alternating multilayers of ultrathin 211 and 123. Microscopy studies in Figs 1 and 2 reveal a near-uniform dispersion of second phase 211 nanoparticles grown by the island-growth mode in YBCO, demonstrating consistent epitaxial growth of YBCO around the 211 islands for up to 200 bilayers. The island-growth mechanism is expected for the (211/123) $\times$ N multilayer system, where N is the number of bilayers, because the lattice mismatch of 211 with respect to 123 is of the order of 2%–7%, depending on the growth orientation of 211 with respect to 123. For one film analysed with electron selected area diffraction, the growth orientation was 211 b axis//123 c axis, and the lattice mismatches calculated from reference patterns were, for lattice constants a , b and c : $a_{211}/2 \times (a,b)_{\text{avg}123} = -7.4\%$, where $(a,b)_{\text{avg}123}$ is the average of a and b in 123; $c_{211, \text{rip-45}^\circ}/(a,b)_{\text{avg}123} = +4.5\%$, where $c_{211, \text{rip-45}^\circ}$ is c rotated 45° in the plane; and $b_{211}/c_{123} = +4.5\%$. Other growth orientations might be possible, but they have not yet been observed. A large lattice mismatch favours the growth of island nanoparticles, because deposition is preferred on energetically favourable island-nucleated phases rather than the lattice-mismatched substrate or base layer¹⁷.

As shown in Fig. 2, the growth of (211-nanoparticle/123-film) $\times$ N multilayer films produced a superlattice-type structure with the 211 layers deposited as nanoparticles rather than homogeneous epitaxial layers. A repeating (nanoparticle/film) structure has been achieved with semiconductor materials that are described as three-dimensional quantum-dot or binary superlattices or crystals^{18–20}; however, to our knowledge, this type of structure has only been recently explored by our group in superconductors with initial experiments testing the (211/123) $\times$ N structure with N bilayers up to 100 (ref. 21). An advantage of the (211/123) $\times$ N structure for flux pinning is that it is possible to have some degree of control over the nanoparticle dispersion parameters, for example, by changing the 211 deposition conditions and varying the 123 layer thickness. These films can also provide a good experimental structure to test theories of flux pinning in high-temperature type II superconductors for uniform dispersions of point-like defects²².

Thin film samples were prepared by pulsed laser deposition (PLD) under conditions described previously²¹. Layers of 123 and 211 nanoparticles were deposited by ablation of separate 123 and 211 composition targets, using parameters chosen to optimize superconducting properties of the 123 phase²³. The O $_2$ deposition pressure was 300 mTorr. Unless noted otherwise, the composite films were deposited by computer program control of the laser and automated control of target movement. A delay of about 13 s was used between 123 and 211 depositions, during which the deposition was stopped and the targets were changed and moved into position. The plumes for 211 and 123 depositions were similar in size and shape, which allowed the use of the same PLD parameters. Substrates were LaAlO $_3$ or SrTiO $_3$ single crystals. The 123 layer and the 211 ‘pseudo-layer’ thicknesses were estimated by calibrating the deposition rates of both 123 and 211 for many depositions. These rates and the total sample thickness were used to calculate the

thicknesses of the 211 ‘pseudo-layer’ and 123 layer, assuming continuous coverage.

The formation of nanoparticles on the surfaces of films are shown in Fig. 1, as imaged with scanning electron microscopy (SEM). The nanoparticles are assumed to be insulating 211 phase, because the contrast seen for the particles is probably due to the insulating properties which cause charging and increased secondary electron emission. The nanoparticles were always observed on the film surfaces, with sizes and densities varying depending on 211 deposition parameters; typical particle sizes were 15 nm and particle densities about 1 – 1.3×10^{11} particles cm $^{-2}$, equivalent to a fluxon field of about 2–3 T, as shown in Fig. 1. Nanoparticles were observed to have diameters about twice as large on the surface as on the inside (as imaged with transmission electron microscopy (TEM) cross-sections). Assuming similar 211 area coverage for the surface and inside layers, the 211 densities inside the film are conservatively estimated as about four times higher, or $\sim(4\text{--}5) \times 10^{11}$ particles cm $^{-2}$. The actual number density may be higher, however, if the nanoparticle heights on the surface layers are also larger, which would increase the 211 volume. A similar calculation, using particle sizes and assuming the 211 layer achieves full coverage based on the calibrated deposition rate, indicates that the 211 densities might be as high as $(7\text{--}8) \times 10^{11}$ particles cm $^{-2}$. Therefore an initial estimate of the nanoparticle densities is $>4 \times 10^{11}$ particles cm $^{-2}$.

The increase of nanoparticle size on the surface also suggests the holding time at temperatures $>700^\circ\text{C}$ of the surface layer affects the 211 nucleation density and particle size, presumably by mechanisms of coalescence and ripening²⁰. Note that particle formation was especially enhanced on plateau edges and corners, which are expected to be energetically more favourable for coalescence sites¹⁷.

The formation of 211 nanoparticles was observed with cross-sectional TEM, as shown in Fig. 2. An important result of the TEM studies was that YBCO maintains a coherent structure as it grows around the 211 material. The uniform layering of the 211 phase can be clearly seen in Fig. 2a, which continued up to the surface of the $\sim 2\text{-}\mu\text{m}$ -thick film. The particle size and distribution are more clearly discerned in the bright-field transmission images in Fig. 2b

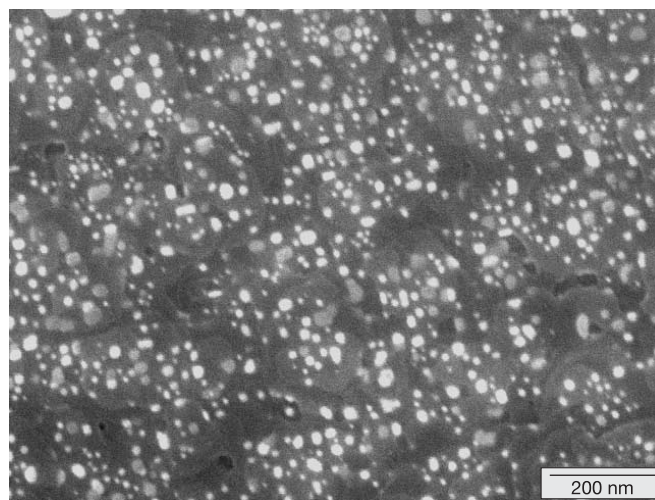


Figure 1 Micrograph of a YBCO + nanoparticle film showing surface nanoparticle formation. The image is from a (211 ~ 1.1 nm/123 ~ 12.6 nm) $\times$ 25 multilayer film. White-image nanoparticles have average particle size 14.8 ± 0.7 nm and areal number density $\sim 1.1 \times 10^{11}$ particles cm $^{-2}$. Note that formation of nanoparticles occurred preferentially on the edges or corners of 123 ledges. Imaging was done with an SEM with 5 kV through a lens secondary detector in high-resolution mode, and imaging depth about 3 nm.

and c for different films. However, we note that TEM images provide a two-dimensional representation of particles embedded throughout the three-dimensional foil volume, and so the three-dimensional dispersion of particles is not yet precisely known. Similarly, the measured two-dimensional areal number density of nanoparticles for the films in Figs 2 and 3 ($\sim 5\text{--}6 \times 10^{11}$ particles cm^{-2}) cannot be used to calculate actual three-dimensional volumic or two-dimensional areal densities without knowing the precise foil thickness. For films in this study, the foil thickness was not measured at this time. The 211 particles are cross-sections of rectangular shape, surrounded by darker areas which can be considered a localized stress field that surrounds the particles²⁴.

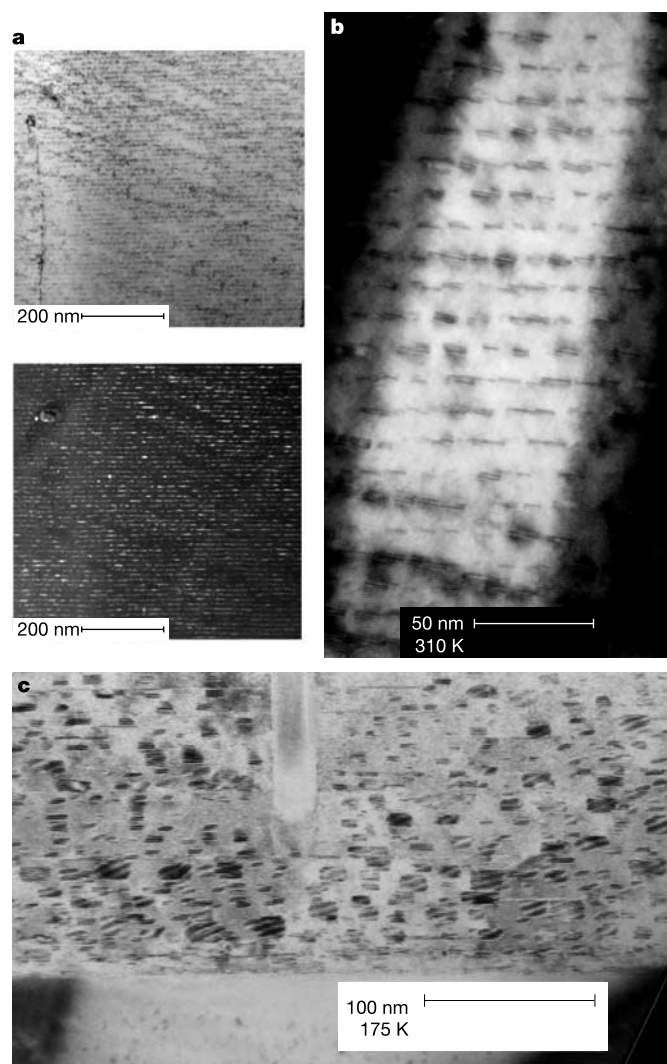


Figure 2 Transmission electron micrographs of YBCO + nanoparticle films, showing repeat layering structure and nanoparticle formations. Images are cross-sections of different films: **a**, **b**, (211 ~ 0.9 nm/123 ~ 10.4 nm) $\times 200$; **c**, (211 ~ 1.7 nm/123 ~ 6.6 nm) $\times 35$. **a**, A bright-field image (top) and a corresponding centred aperture dark-field image (bottom). The dark-field imaging conditions in **a** were established with a diffraction vector from the particle phase, thereby allowing the particle distribution to show up clearly against the 123 matrix phase. Nanoparticles are rectangular-shaped, about 8 nm in width in **b**, and 10–20 nm in width in **c**. Note the dislocation line in **a** and stress fields surrounding the 211 defects, and especially the formation of columnar-type defects in **b** and **c**, showing vertically stacked 211 connected by localized stress regions. The >100 -nm-high vertically aligned defect is probably *a*-axis-oriented 123, which was only rarely observed in the multilayer films. The TEM for microscopy studies was a Philips CM-200 with a field-emission gun source.

These localized stress regions increase the size of the non-superconducting volume surrounding the defect, and thereby change the pinning properties.

As observed, especially in Fig. 2c, for a film with a thinner 123 layer spacing of ~ 6.5 nm, the localized stress region can extend between the nanoparticles that are vertically aligned, creating pseudo-columnar defects. For the sample in Fig. 2c, several growth mechanisms might have caused the more enhanced vertical alignment: (1) the holding time between layers was achieved by manually changing the target position with a longer average holding time of about 80 s, compared to 13 s achieved with automated target changeover, and (2) the film had a very low 123 layer thickness of about 6.5 nm. Vertical alignment of nanoparticles in superlattice-type structures is affected by both the layer spacing and the holding time²⁰, with a longer holding time presumably enhancing vertical alignment by allowing the island particles more time to nucleate on energetically disturbed sites. So it is not unexpected that some amount of vertical alignment of particles was observed in these films, but not in all cases. The formation of columnar defects is of interest, as columnar defects are expected to have greater pinning strength than do point defects^{3,4}.

Selected area diffraction patterns confirmed the presence of two distinct phases, consistent with identification as 123 and 211 with orthorhombic structure and slightly varying lattice parameters or similar phase. The selected area diffraction peaks were shifted slightly compared to reference 123 and 211 patterns, suggesting the phases are in relative states of stress or distorted from lattice strains caused by the lattice parameter mismatch. The 123 phase is highly *c*-axis oriented in the superlattice films²¹ and the in-plane orientation of 123 as measured by Φ scans is in general excellent with a full-width half-maximum (FWHM) of about 1.2° .

The effect of the 211 nanoparticle addition on critical current densities is shown in Fig. 3. Transport J_c (J_{ct}) for the composite films are compared to 123 films fabricated by PLD and the BaF_2 process^{12,13}. In particular for applied magnetic field greater than

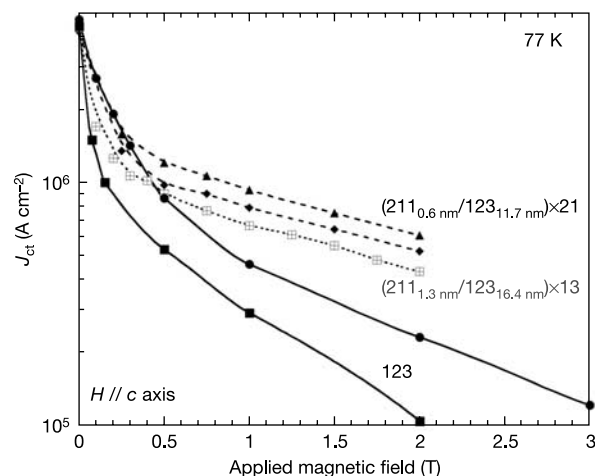


Figure 3 Critical current density as a function of applied magnetic field for YBCO + nanoparticle films compared to pure YBCO films. 123 films (solid lines) were deposited by PLD (filled squares)¹², and recent sample data provided by Oak Ridge National Laboratory of 123 on CeO_2 -coated $\text{Zr}(\sim 9\%\text{Y})\text{O}_2$ substrates (circles) using the BaF_2 process¹³. Multilayer films (dashed lines) were measured for different 211 and 123 parameters, and for samples from the same deposition run (triangles and diamonds). J_{ct} values were measured on bridge-patterned samples by a four-probe transport method with a $1 \mu\text{V cm}^{-1}$ criterion; bridge dimensions were 0.05–0.075 cm wide and 0.3–0.6 cm in length. J_c was calculated using the total film thickness including 211 and 123 layers. The film thickness, bridge widths and dimensions were measured several times to reduce the standard errors of each of these measurements to $<5\%$.

0.3 T, the composite films had a flatter dependence on applied field, and the absolute values of J_c began to increase with respect to 123 films. The self-field J_{ct} of the composite films were increased to above 4 MA cm^{-2} from initial results of $2\text{--}3 \text{ MA cm}^{-2}$ (ref. 21). The increase of the self-field J_c was achieved by decreasing the 211 particle deposition time and hence the 211 'pseudo' layer thickness from $1\text{--}1.5 \text{ nm}$ to 0.5 nm . A decrease of 211 'pseudo-layer' thickness from 1.0 nm to 0.5 nm increases the volume percentage of the 123 phase and presumably reduces the intrinsic stresses of the 211/123 composites. This decrease in thickness of 211 also increased the T_c from $88.9 \pm 0.2 \text{ K}$ to $90.2 \pm 0.4 \text{ K}$ on average (for a layer thickness of $\sim 11 \text{ nm}$ for 123). These factors combined potentially enhance the zero-field J_c . □

Received 22 April; accepted 28 June 2004; doi:10.1038/nature02792.

1. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Wu, M. K. *et al.* Superconductivity at 93 K in a new mixed-phase Y-Ba-Cu-O compound system at ambient pressure. *Phys. Rev. Lett.* **58**, 908–910 (1987).
3. Larbalestier, D., Gurevich, A., Matthew Feldmann, D. & Polyanskiy, A. High- T_c superconducting materials for electric power applications. *Nature* **414**, 368–377 (2001).
4. Matsushita, T. Flux pinning in superconducting 123 materials. *Supercond. Sci. Technol.* **13**, 730–737 (2000).
5. Cival, L. *et al.* Vortex confinement by columnar defects in $\text{YBa}_2\text{Cu}_3\text{O}_7$ crystals: enhanced pinning at high fields and temperatures. *Phys. Rev. Lett.* **67**, 648–651 (1991).
6. Selvamankam, V. *et al.* Fabrication of 100 A class, 1 m long coated conductor tapes by metal organic chemical vapor deposition and pulsed laser deposition. *Physica C* **392–396**, 859–862 (2003).
7. Verebelyi, D. T. *et al.* Uniform performance of continuously processed MOD-YBCO-coated conductors using a textured Ni-W substrate. *Supercond. Sci. Technol.* **16**, L19–L22 (2003).
8. Groves, J. R. *et al.* Recent progress in continuously processed IBAD MgO template meters for HTS applications. *Physica C* **382**, 43–47 (2002).
9. Goyal, A. *et al.* Recent progress in the fabrication of high- J_c tapes by epitaxial deposition of YBCO on RABiTS. *Physica C* **357–360**, 903–913 (2001).
10. Balachandran, U. *et al.* Development of coated conductors by inclined substrate deposition. *Physica C* **392–396**, 806–814 (2003).
11. Kakimoto, K., Iijima, Y. & Saitoh, T. Fabrication of long-Y123 coated conductors by a combination of IBAD and PLD. *Physica C* **392–396**, 783–789 (2003).
12. Yamasaki, H., Nakagawa, Y., Sawa, A., Obara, H. & Develos, K. Flux pinning effects of twin boundaries studied with unidirectionally twinned YBCO films. *Physica C* **372–376**, 1885–1889 (2002).
13. Feenstra, R., Christen, D. K., Klabunde, C. E. & Budai, J. D. Role of oxygen vacancies in the flux-pinning mechanism, and hole-doping lattice disorder in high-current-density $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ films. *Phys. Rev. B* **45**, 7555–7558 (1993).
14. Larbalestier, D. C. & Maley, M. P. Conductors from superconductors: conventional low-temperature and new high-temperature superconducting conductors. *MRS Bull.* **18**, 50–56 (Aug. 1993).
15. Dam, B. *et al.* Origin of high critical currents in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ superconducting thin films. *Nature* **399**, 439–442 (1999).
16. Pan, V. M. *et al.* Nature of magnetic field and angular dependencies of the critical current density in epitaxial HTS $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ films. *Physica C* **388–389**, 431–432 (2003).
17. Reichelt, K. Nucleation and growth of thin films. *Vacuum* **38**, 1083–1099 (1988).
18. Redl, F. X., Murray, K.-S., Cho, C. B. & O'Brien, S. Three-dimensional binary superlattices of magnetic nanocrystals and semiconductor quantum dots. *Nature* **423**, 968–971 (2003).
19. Springholz, G., Holy, V., Pinczolis, M. & Bauer, G. Self-organized growth of three-dimensional quantum-dot crystals with FCC-like stacking and a tunable lattice constant. *Science* **282**, 734–737 (1998).
20. Liu, P., Zhang, Y. W. & Lu, C. Self-organized growth of three-dimensional quantum-dot superlattices. *Appl. Phys. Lett.* **80**, 3910–3912 (2002).
21. Haugan, T. *et al.* Island-growth of Y_2BaCuO_5 nanoparticles in $(211\text{--}1.5 \text{ nm}/123\text{--}10 \text{ nm})\text{xN}$ composite multilayer structures to enhance flux pinning of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films. *J. Mater. Res.* **18**, 2618–2623 (2003).
22. Gross, R. E. & Campbell, A. M. Numerical calculation of elastic pinning parameters by point pins. *Physica C* **260**, 188–196 (1996).
23. Haugan, T., Barnes, P. N., Brunke, L., Maartense, I. & Murphy, J. Effect of O_2 partial pressure on $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ thin film growth by pulsed laser deposition. *Physica C* **297**, 47–57 (2003).
24. Zhu, Y., Cai, Z. X., Budhani, R. C., Suenaga, M. & Welch, D. O. Structures and effects of radiation damage in cuprate superconductors irradiated with several-hundred-MeV heavy ions. *Phys. Rev. B* **48**, 6436–6450 (1993).

Acknowledgements The Air Force Office of Scientific Research supported this work. We thank J. Murphy, L. Brunke, J. Evans and T. Campbell for experimental assistance, and S. Apt of UES Inc. at the Wright-Patterson AFB Materials Directorate for assistance with SEM and TEM. We also thank R. Feenstra and A. A. Gapud at Oak Ridge National Laboratory (ORNL) for providing $J_c(H)$ data for a reference 123 film.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J.H. (timothy.haugan@wpafb.af.mil).

Direct evidence for atomic defects in graphene layers

Ayako Hashimoto¹, Kazu Suenaga¹, Alexandre Gloter^{1,2}, Koki Urita^{1,3} & Sumio Iijima¹

¹Research Center for Advanced Carbon Materials, National Institute for Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8565, Japan

²Laboratoire de Physique des Solides, CNRS UMR 8502, Université Paris-Sud, Orsay, 91405, France

³Graduate School of Science and Technology, Chiba University, Chiba, 263-8522 Japan

Atomic-scale defects in graphene layers alter the physical and chemical properties of carbon nanostructures^{1,2}. Theoretical predictions have recently shown that energetic particles such as electrons and ions can induce polymorphic atomic defects in graphene layers as a result of knock-on atom displacements^{3,4}. However, the number of experimental reports on these defects is limited^{5,6}. The graphite network in single-walled carbon nanotubes has been visualized by transmission electron microscopy (TEM) and their chiral indices have been determined^{7,8}. But the methods used require a long image acquisition time and intensive numerical treatments after observations to find an 'average' image, which prevents the accurate detection and investigation of defect structures. Here we report observations *in situ* of defect formation in single graphene layers by high-resolution TEM. The observed structures are expected to be of use when engineering the properties of carbon nanostructures for specific device applications.

Pentagon–heptagon pairs⁹, mono-vacancies and multi-vacancies^{4,10–13} and adatoms^{3,14} are typical stable graphene defects that have been predicted. (The metastable interstitial-vacancy defect described in ref. 13 is between two graphene planes and is therefore not exactly the same defect as that investigated here.)

We fabricated a desired contrast transfer function (CTF) that enables selective visualization of the graphitic network (see Methods and Supplementary Figs S1 and S2). Under this experimental condition, each zig-zag chain should turn up as a dark line in TEM images and a bright spot should appear in the middle of each hexagonal carbon ring (see Fig. 1a and inset).

In a perfect single-walled carbon nanotube (SWNT; Fig. 1a, b), the moiré pattern formed by the rolled-up graphene layer is clearly visible between two intense dark lines corresponding to the vertical tube walls. By using optical diffraction (Fourier transform), two symmetric hexagons are identified on the tube axis (Fig. 1b inset). Each hexagon in Fourier space corresponds to a single graphene layer (front or back of the SWNT) and represents the orientation of the zig-zag chain as illustrated in Fig. 1a (inset) to the tube axis, from which the apparent chiral angle can easily be deduced¹⁵. The apparent chiral angle and the tube diameter measured in the high-resolution TEM (HR-TEM) image provide one or at most three solution(s) of the chiral index (n , m) for the nanotube under examination. Consequently, the chiral index of SWNT has been assigned as (13, 8) (Fig. 1c) (or (13, 7) or (12, 8) with pessimistic error bars).

Such an analysis can be performed even on a SWNT of mixed chirality (Fig. 1d). This non-uniform SWNT was produced by electron bombardment, as described in previous^{16,17} reports of processes used to thin the walls of SWNTs. Optical diffraction analysis of the SWNT tells us that the tube with an axial heterojunction is assigned to (17, 0) at the upper part and (18, 0) at the lower part. It implies that this serial junction created *in situ* here consists of a semiconducting (17, 0) SWNT and a metallic (18, 0) SWNT, and then may generate a nanodiode. A pentagon–heptagon pair is most probably responsible for the serial junction of two parts of SWNTs with different chiral indices. An image simulation of an

adjacent pentagon and heptagon (depicted in Fig. 1e, f) fits quite well with the defective region in the observed image. Direct imaging of the graphitic network allows the local analysis of chiral index of composite nanotubes, which other assignment techniques such as Raman or fluorescent spectroscopy are unable to provide.

Such a topological defect can be seen more clearly in a plan view of graphene layers. Next we show the topological defects induced in a single graphene layer observed *in situ* by HR-TEM. Figure 2a shows a plan view of a graphene layer in a carbon nanostructure with a large diameter¹⁸. After several tens of seconds of irradiation with an electron beam, a missing row of zig-zag chain was clearly observed in the graphene network (compare Fig. 2a (before irradiation) with Fig. 2b (after irradiation)). Such a basal plane dislocation in the graphene layer can be explained by the existence of a topological defect². The pentagon–heptagon pair is one of the reasonable defect models from the viewpoint of energetic stability and can be induced in a graphene layer through the Stone–Wales (SW) process as illustrated (Fig. 2c)⁹. Shown in Fig. 2d is a simulated HR-TEM image of the graphite network with a pentagon–heptagon pair, which agrees quite well with the experimentally observed image. A large cluster of up to 500 atoms has been built for the HR-TEM image simulations. Defect structures were first relaxed by using a semi-empirical potential. In the vicinity of the defect we

used the *ab initio* relaxed structure within this large cluster in the literature². The rows of zig-zag chains in three directions cannot all be seen in the HR-TEM images (Fig. 2a, b) because the graphene layer is largely inclined against the incident beam beyond the limit of the inclination angle ($\sim 20^\circ$ in this experiment), where the projected spacing of the graphite network goes off the range of the targeted CTF peak. The extent of the local strain field, as defined by loss of contrast in the HR-TEM images, is estimated at about 1.2 nm. Qualitatively, this fits well with the HR-TEM simulation based on the relaxed pentagon–heptagon structure showing a blurred area of 1 or 2 nm around the dislocation core (Fig. 2d). During the formation of this topological defect, an elastic deformation or a bending of the graphene layer can be observed in the time-sequential HR-TEM images. This proves that the SW transformation should have a key role in nanotube relaxation under strain, as previously suggested by a theoretical prediction¹⁹.

Imaging a point defect, such as an atomic vacancy, on a graphene layer is even more challenging and is of crucial importance because it refers directly to the physical and even the magnetic properties of this material^{14,20}. Until now, only scanning probe microscopy has been able to investigate the structure of such defects at the scale of an individual object or of a sparse collection of objects^{21,22}. Figure 3a, b shows two HR-TEM images of a unique graphene layer, extracted at $t = 0$ s and $t = 320$ s from a series of hundreds of images. It is evident that several white spots have appeared under electron irradiation, and more than six spots could undoubtedly be isolated from the noise level, leading to a crude estimation of about 0.3 point defects per nm^2 in a graphene layer (see Supplementary Movie 1). The cross-section estimated from this experiment is 160 barn ($160 \times 10^{-10} \text{ nm}^2$), which is quite close to a theoretical value found in ref. 20 for an atom displacement (180 barn). The three

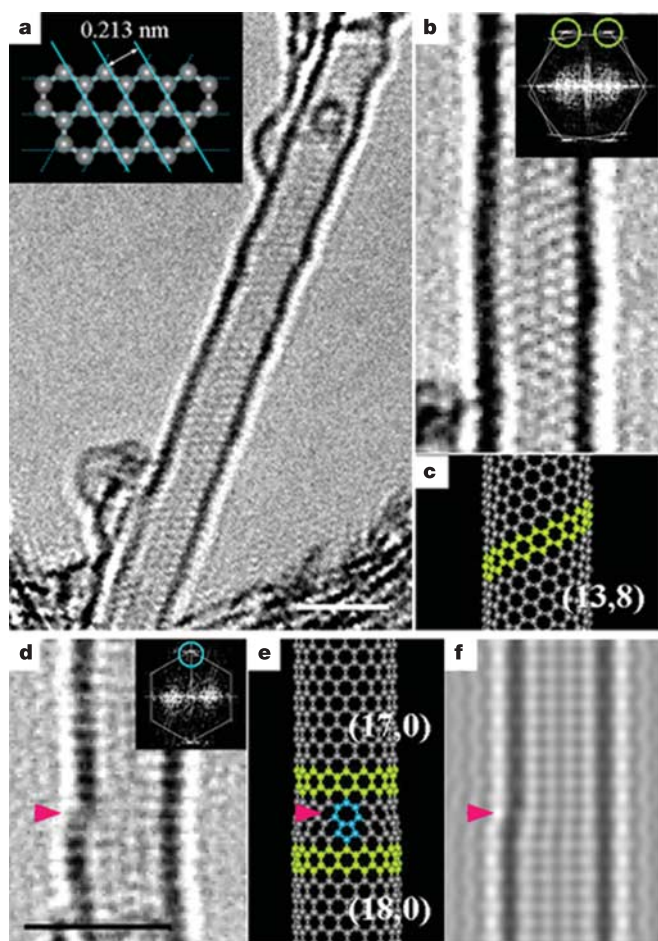


Figure 1 Chiral index determination of SWNTs. **a**, **b**, Chiral index determination of a SWNT. **a**, A typical HRTEM image of a SWNT with enhanced contrast of the zig-zag chain (inset). **b**, Moiré pattern and its optical diffraction (inset). **c**, A best-fit model of the SWNT with the determined chiral index (13, 8). **d**, A cross-sectional view of a topological defect at the side of a SWNT. **e**, A pentagon–heptagon pair is presumably responsible for the defect structure and generates a serial junction of two zig-zag nanotubes (17, 0) and (18, 0). **f**, A simulated image for the SWNT with the defect rotated by 90° . Scale bar, 2 nm.

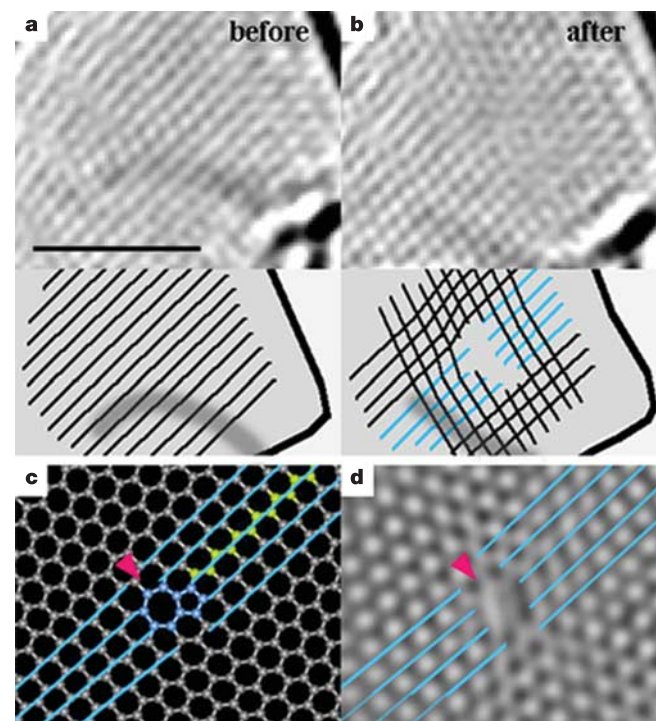


Figure 2 *In situ* observation of a topological defect induced in a graphene layer. **a**, **b**, HR-TEM images of a single graphene layer with a topological defect induced by electron irradiation (before (**a**) and after (**b**)). An edge dislocation is unambiguously visible at the middle of the network where one zig-zag chain is missing through it. The missing zig-zag chain is shown schematically in the bottom part of each panel. **c**, An atomistic model of the pentagon–heptagon pair in the graphitic network. **d**, A simulated HR-TEM image shows a good comparison with the HR-TEM image shown in **b**. Scale bar, 2 nm.

neighbouring defects are presented with increased contrast as an inset to Fig. 3b. The intensity profiles are also presented at the position of these three defects in both images for the initial and final states (Fig. 3a, b, inset). In the initial state, the profile shows only the modulation due to the zig-zag chains, with a periodicity close to 0.213 nm. After irradiation, this modulation is more difficult to perceive because of the lattice disorder arising from the point defects²³. In contrast, these defects appear strongly within the intensity line. HR-TEM simulations have been performed to predict the contrast arising for the most expected point defects on a single graphene layer. In Fig. 3c, d the HR-TEM contrast for one carbon atom vacancy (V_1), two neighbouring vacancies (or an intraplanar di-vacancy, V_2), a pentagon–octagon–pentagon (5–8–5) defect and a carbon adatom can be seen. In Fig. 3c, the atomic positions of the carbon atoms in the vicinity of the mono-vacancies and di-vacancies have not been reconstructed. Nevertheless, theoretical studies^{11,24} have predicted that intraplanar relaxation is weak for a mono-vacancy and, indeed, no difference has been seen in our HR-TEM simulation with the use of either relaxed or unrelaxed atomic positions. In contrast, an interplanar di-vacancy can strongly reconstruct and transform to an agglomeration of non-hexagonal rings such as the 5–8–5 defect shown (Fig. 3c). The adatom has also been positioned in the theoretically predicted equilibrium structure—that is, to be in a bridge-like structure—between two surface atoms^{3,14}. The major issue of the HR-TEM simulations is that, even for a mono-vacancy, the contrast might be strongly enhanced at the centre of the neighbouring hexagons. This ‘delocalization’ effect is particularly strong at the chosen defocus. The three point defects imaged (Fig. 3b) can then be simulated by the removal of only three carbon atoms from a graphene layer as shown (Fig. 3e) and we were unable to obtain a better match by using various other types of (multi-)vacancies or non-hexagonal clusters. This demonstrates the sensitivity of HR-TEM to a single carbon vacancy and strongly suggests that the mono-vacancy does indeed exist in carbon nanostructures for a macroscopic time under

this experimental condition (at least a few seconds) even though the expected energy for vacancy formation in the graphene layer (7.0 eV in ref. 20) is rather high compared with most metals. These vacancies are immobile and do not merge together during observation. A continuous irradiation leads to a collapse of the nanotube so that the strain energy due to the vacancy formation can be relaxed.

Another important issue of the present study is to identify the carbon adatom on the graphene layer. Figure 4a shows a sequence of HR-TEM images recorded *in situ* on one of the SWNTs. Several dark spots and bright spots frequently appear and disappear on the SWNT during observation before it collapses (see also Supplementary Movies 2 and 3), whereas the bright spots correspond to the vacancies as described above, the dark spots should stand for carbon atoms that were knocked-on and then attached to the nanotube surface. However, although these adatoms often move around and disappear during the observation, they are indeed stable for a few seconds on average during several shutter periods. Adatoms have been predicted to be mobile and to have a lower diffusion barrier (0.47 eV)¹⁴ than that for vacancy migration (1.7 eV)¹¹. Note that the adatoms appear mostly in the vicinity of the vacancies (see also Supplementary Movie 4), because this combination of vacancy and neighbouring adatom has been predicted to be metastable and long lived^{4,25,26}. This is direct proof of a recombination barrier between vacancies and adatoms, which are mostly believed to remerge instantaneously when they come close to each other. In addition, more stabilized carbon adatoms on curved graphene layers are pointed out in ref. 12 and the mobility of the adatoms can be decreased in the vicinity of vacancies because of a possible off-plane relaxation, for instance.

To assign these defective structures more precisely, systematic HR-TEM image simulations have been performed for various combinations of adatom(s) and (multi-) vacancy. A best fit with one of the experimental images (Fig. 4b) extracted from Supplementary Movie 4 is shown in Fig. 4c. This structural model consists of three adatoms and one di-vacancy in a planar graphene layer. It again implies adatom mobility because the defect cannot be created locally by one (or two) knock-out events. Two contrast profiles derived from the simulated image show quantitative agreement with the experimental profiles recorded at the Scherzer defocus (Fig. 4d). Putting heavier atoms such as Fe mono-atoms

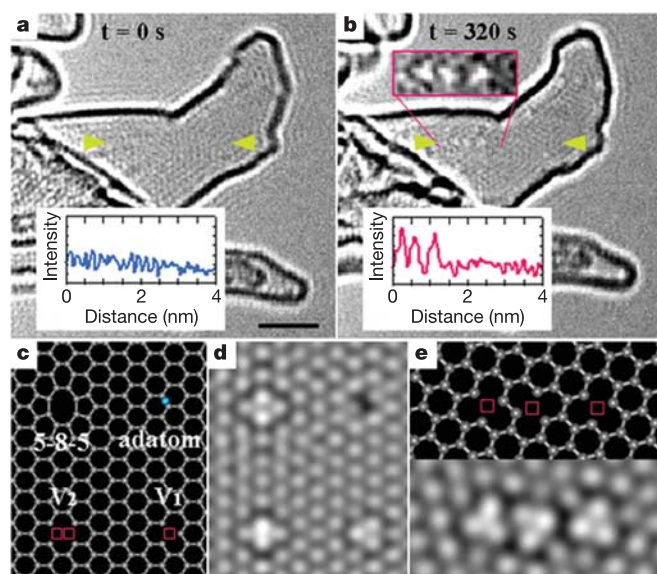


Figure 3 *In situ* observation of vacancy formation in a graphene layer. **a**, **b**, HR-TEM images of a single-walled carbon nanotube recorded before electron irradiation (**a**) and after 320 s of irradiation (**b**). Intensity profiles (insets) between the two arrows illustrate the contrast measured at the carbon vacancies. **c**, **d**, Atomistic models (**c**) and simulated HR-TEM image (**d**) for a graphene layer with an adatom, unrelaxed vacancies (V_1 and V_2) and a 5–8–5 rearrangement. Squares indicate the location of the vacancy. **e**, Atomic model and simulated HR-TEM image with three mono-vacancies, giving the best match to the experimentally observed defect structures (**b**, inset). Scale bar, 2 nm.

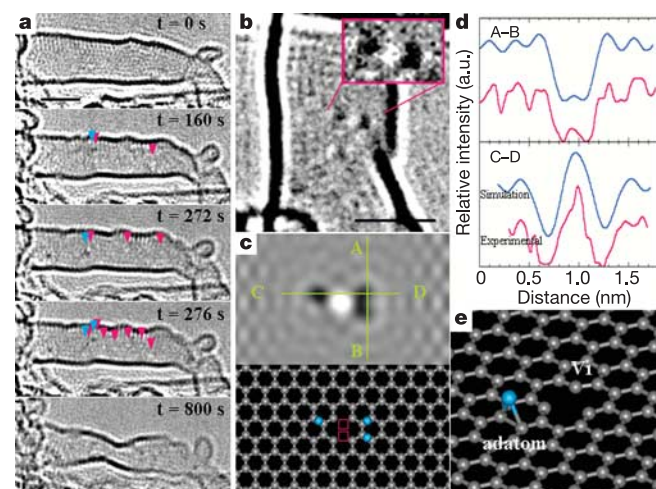


Figure 4 *In situ* observation of adatom–vacancy pair formation. **a**, HR-TEM images of a SWNT during electron irradiation (the total acquisition time was ~800 s). The bright spots correspond to the vacancies (red arrows) and the dark to adatoms (blue arrows). **b**, An extracted image from Supplementary Movie 4. **c**, Simulation for three adatoms and di-vacancy model (not energetically relaxed) fits quite well with the experiment. **d**, Two contrast profiles derived from the simulation and the experimental image show quantitative agreement. **e**, The most prolific defect model with an adatom–vacancy pair (see also Supplementary Movie 2). Scale bar, 2 nm.

into the simulation (representing possible impurities from a catalyst) instead of carbon adatoms gives a considerable deviation from the experimental profiles, which supports again the notion that the defect structure involves only carbon adatoms and vacancies. The more frequent defect structure of an adatom–vacancy pair is shown schematically in Fig. 4e.

Defects in a graphene layer, such as topological defects, vacancies and adatoms, have been experimentally proved to be numerous and stable under electron irradiation. We can envisage more diversified applications in nanocarbon materials by taking advantage of these defects, which can be induced locally during electron irradiation. Combining the present experiments with the scanning probe microscopy technique and its associated electron transport measurements^{27,28} is of particular interest and might allow us to obtain more comprehensive information on the physical and chemical properties of defective carbon nanostructures. □

Methods

Electron microscopy

A Schottky-type field emission gun (JEOL, JEM-2010F) was used for HR-TEM. We employed optimized parameters for imaging, well fitted to a high-efficiency charge-coupled device (CCD)-based camera (Gatan model 794) with a fibre-optic coupling to the YAG scintillators. To enhance phase contrast and decrease thermally induced drift in the specimen as much as possible, a multi-exposure procedure was employed so that we could decrease the exposure time to 1 s and sum a few images with specimen-drift corrections to limit electron irradiation damage and the consequent thermally induced structural changes, high-speed blanking was used to prevent exposure of the specimen to the electron beam during each readout by the CCD detector. A typical electron dose was $\sim 60,000$ electrons nm^{-2} for a HR-TEM image. To realize both sufficient sensitivity and sufficient resolution to visualize the single graphene layer, we used a lower accelerating voltage (120 keV) for the incident electron beam close to the displacement threshold, for the following three reasons: first, to double the scattering cross-section of carbon atoms being observed; second, so that carbon adatoms were not blasted away from the graphene layers and could therefore be observed (note that the threshold for electron irradiation on carbon materials is taken to be 80–140 keV; and last, to optimize the efficiency of the CCD detector for image acquisition. As indicated by the red graph in Supplementary Fig. S1 (bottom inset), we carefully chose the experimental conditions as a compromise with the moderate accelerating voltage of the incident electron beam so that the CTF could have a local maximum at 0.21–0.23 nm, which corresponds exactly to the repeat distance of the zig-zag chain of a graphene layer (Supplementary Fig. S1, top inset). The confidence level for the detection of a single carbon atom exceeds 80%. All the experiments shown here were performed at room temperature because raising the temperature might have led to an instantaneous relaxation of these atomic defects.

Chiral index determination

The apparent chiral angle (α) can be measured to an accuracy of $\sim 3^\circ$; measurement of the diameter (d) in the HR-TEM image might involve about 10% error, as previously quoted in refs 15 and 29. We made a systematic study with a series of image simulations for the nanotubes of various diameters (Supplementary Fig. S3). Taking all of the microscope parameters into account, the true diameter can be therefore deduced from the apparent tube diameter to within the 3% error. The inclination of the specimen to the incident electron beam cannot be neglected in this experiment. We selected the tubes to examine which were as parallel as possible to the observed plane within a few degrees of inclination. Optical diffraction is strongly influenced by the inclination of the sample, whereas conventional electron diffraction is weakly dependent on it. The distortion and disappearance of symmetrical hexagons in optical diffraction is able to assist in the estimation of inclination angles. Our simulation tells us that the chiral angle can be measured to an accuracy of $\sim 3^\circ$ when the inclination angle is within 10° .

Received 13 February; accepted 6 July 2004; doi:10.1038/nature02817.

- Hansson, A., Paulsson, M. & Stafström, S. Effect of bending and vacancies on the conductance of carbon nanotubes. *Phys. Rev. B* **62**, 7639–7644 (2000).
- Ewels, C. P., Heggge, M. I. & Briddon, P. R. Adatoms and nanoengineering of carbon. *Chem. Phys. Lett.* **351**, 178–182 (2002).
- Nordlund, K., Keinonen, J. & Mattila, T. Formation of ion irradiation induced small-scale defects on graphite surfaces. *Phys. Rev. Lett.* **77**, 699–702 (1996).
- Krasheninnikov, A. V., Nordlund, K., Sirviö, M., Salonen, E. & Keinonen, J. Formation of ion-irradiation-induced atomic-scale defects on walls of carbon nanotubes. *Phys. Rev. B* **63**, 245405 (2001).
- Kelly, K. F. & Halas, N. J. Determination of α and β site defects on graphite using C_{60} -adsorbed STM tips. *Surf. Sci.* **416**, L1085–L1089 (1998).
- Ouyang, M., Huang, J.-L., Cheung, C. L. & Lieber, C. M. Atomically resolved single-walled carbon nanotube intramolecular junctions. *Science* **291**, 97–100 (2001).
- Meyer, R. R. *et al.* A composite method for the determination of the chirality of single walled carbon nanotube. *J. Microsc.* **212**, 152–157 (2003).
- Zuo, J. M., Vartanyants, I., Gao, M., Zhang, R. & Nagahara, L. A. Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science* **300**, 1419–1421 (2003).
- Stone, A. J. & Wales, D. J. Theoretical studies of icosahedral C_{60} and some related species. *Chem. Phys. Lett.* **128**, 501–503 (1986).

- Telling, R. H., Ewels, C. P., El-Barbary, A. A. & Heggge, M. I. Wigner defects bridge the graphite gap. *Nature Mater.* **2**, 333–337 (2003).
- El-Barbary, A. A., Telling, R. H., Ewels, C. P., Heggge, M. I. & Briddon, P. R. Structure and energetics of the vacancy in graphite. *Phys. Rev. B* **68**, 144107 (2003).
- Krasheninnikov, A. V. *et al.* Adsorption and migration of carbon adatoms on carbon nanotubes: Density-functional *ab initio* and tight-binding studies. *Phys. Rev. B* **69**, 073402 (2004).
- Ewels, C. P., Telling, R. H., El-Barbary, A. A., Heggge, M. I. & Briddon, P. R. Metastable Frenkel pair defect in graphite: Source of Wigner energy? *Phys. Rev. Lett.* **91**, 25505 (2003).
- Lehtinen, P. O. *et al.* Magnetic properties and diffusion of adatoms on a graphene sheet. *Phys. Rev. Lett.* **91**, 17202 (2003).
- Kociak, M., Hirahara, K., Suenaga, K. & Iijima, S. How accurate can the determination of chiral indices of carbon nanotubes be? *Eur. Phys. J. B* **32**, 457–469 (2003).
- Ajayan, P. M., Ravikumar, V. & Charlier, J.-C. Surface reconstructions and dimensional changes in single-walled carbon nanotubes. *Phys. Rev. Lett.* **81**, 1437–1440 (1998).
- Smith, B. W. & Luzzi, D. E. Electron irradiation effects in single wall carbon nanotubes. *J. Appl. Phys.* **90**, 3509–3515 (2001).
- Iijima, S. *et al.* Nano-aggregates of single-walled graphitic carbon nano-horns. *Chem. Phys. Lett.* **309**, 165–170 (1999).
- Yakobson, B. I., Samsonidze, G. & Samsonidze, G. G. Atomistic theory of mechanical relaxation in fullerene nanotubes. *Carbon* **38**, 1675–1680 (2000).
- Banhart, F. Irradiation effects in carbon nanostructures. *Rep. Prog. Phys.* **62**, 1181–1221 (1999).
- Hahn, J. R., Kang, H., Song, S. & Jeon, I. C. Observation of charge enhancement induced by graphite atomic vacancy: A comparative STM and AFM study. *Phys. Rev. B* **53**, R1725–R1728 (1996).
- Orlikowski, D., Buongiorno Nardelli, M., Bernholc, J. & Roland, C. Theoretical STM signatures and transport properties of native defects in carbon nanotubes. *Phys. Rev. B* **61**, 14194–14203 (2000).
- Asari, E., Kitajima, M., Nakamura, K. G. & Kawabe, T. Thermal relaxation of ion-irradiation damage in graphite. *Phys. Rev. B* **47**, 11143–11148 (1993).
- Hjort, M. & Stafström, S. Modeling vacancies in graphite via the Hückel method. *Phys. Rev. B* **61**, 14089–14094 (2000).
- Krasheninnikov, A. V., Nordlund, K. & Keinonen, J. Production of defects in supported carbon nanotubes under ion irradiation. *Phys. Rev. B* **65**, 165423 (2002).
- Lu, A. J. & Pan, B. C. Nature of single vacancy in achiral carbon nanotubes. *Phys. Rev. Lett.* **92**, 105504 (2004).
- Wildöer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. Electronic structure of atomically resolved carbon nanotubes. *Nature* **391**, 59–62 (1998).
- Odom, T. W., Huang, J.-L., Kim, P. & Lieber, C. M. Atomic structure and electronic properties of single-walled carbon nanotubes. *Nature* **391**, 62–64 (1998).
- Kociak, M. *et al.* Linking chiral indices and transport properties of double-walled carbon nanotubes. *Phys. Rev. Lett.* **81**, 155501 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Kociak for his instructions on the chiral index determination of SWNTs; C. Ewels and Y. Miyamoto for discussions on defective carbon structures; and M. Yudasaka and S. Bandow for help with specimen preparation. Work on HR-TEM is supported by the NEDO Nano-Carbon Technology project. A.H. was supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S. (suenaga-kazu@aist.go.jp).

Chemical remodelling of cell surfaces in living animals

Jennifer A. Prescher^{1*}, Danielle H. Dube^{1*} & Carolyn R. Bertozzi^{1,2}

¹Department of Chemistry, University of California, Berkeley, California 94720, USA

²Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

*These authors contributed equally to this work

Cell surfaces are endowed with biological functionality designed to mediate extracellular communication. The cell-surface repertoire can be expanded to include abiotic functionality through the biosynthetic introduction of unnatural sugars into cellular glycans, a process termed metabolic oligosaccharide engineering^{1,2}. This technique has been exploited in fundamental studies of glycan-dependent cell–cell and virus–cell interactions^{3–5} and

also provides an avenue for the chemical remodelling of living cells^{6–8}. Unique chemical functional groups can be delivered to cell-surface glycans by metabolism of the corresponding unnatural precursor sugars. These functional groups can then undergo covalent reaction with exogenous agents bearing complementary functionality. The exquisite chemical selectivity required of this process is supplied by the Staudinger ligation of azides and phosphines, a reaction that has been performed on cultured cells without detriment to their physiology^{7,9}. Here we demonstrate that the Staudinger ligation can be executed in living animals, enabling the chemical modification of cells within their native environment. The ability to tag cell-surface glycans *in vivo* may enable therapeutic targeting and non-invasive imaging of changes in glycosylation during disease progression.

Central to the process of chemical cell-surface remodelling is the concept of bio-orthogonality. The pair of functional groups chosen for the cell-surface transformation must be mutually reactive under physiological conditions and, at the same time, remain inert to the biological environment. We previously developed a reaction that meets these criteria—the Staudinger ligation between azides and a specific class of phosphines (Fig. 1a)⁷. The azide is essentially unreactive with biological nucleophiles yet readily condenses with exogenously delivered phosphine reagents. If the phosphine is functionalized as shown in Fig. 1a, its reaction with azides forms a stable amide bond with concomitant release of nitrogen and methanol and conversion to the phosphine oxide.

Azides can be metabolically introduced into cell-surface glycans by virtue of the unnatural substrate tolerance of certain carbohydrate biosynthetic pathways. For example, the unnatural sugar *N*- α -azidoacetylmannosamine (ManNAz) is metabolized by cultured cells to *N*- α -azidoacetyl sialic acid (SiaNAz) and incorporated into membrane glycoconjugates (Fig. 1b)^{7,9}. Azido analogues of *N*-acetylgalactosamine and *N*-acetylglucosamine can be similarly incorporated into cellular glycoproteins^{10,11}. Once displayed on the cell surface, the azides are poised for Staudinger ligation with phosphine compounds to generate unique cell-surface epitopes. In addition, chemical tagging of azide-labelled glycans permits the identification of specific glycoprotein subtypes from the proteome^{10,11}. Beyond its utility in elaborating glycan structures,

the Staudinger ligation has been used in numerous biological applications^{12,13}.

So far, the unique features of the Staudinger ligation have only been exploited in biochemical or cell-based systems. There are obvious advantages, however, to applying this reaction in the physiologically authentic environment of a living organism—especially in the context of metabolic oligosaccharide engineering. Cellular glycosylation patterns could be probed within organs, and cells could be remodelled to change their immunogenic properties in a temporally controlled fashion. Moreover, the ability to chemically alter cell surfaces in a targeted manner is applicable to the diagnosis and treatment of human diseases associated with aberrant glycosylation^{14–16}. To realize this vision would require that azido-sugar metabolism and the Staudinger ligation both proceed in living animals.

Here we report that cells can be chemically modified in laboratory mice by metabolism of peracetylated ManNAz (Ac₄ManNAz) and subsequent Staudinger ligation. We chose the sialic acid biosynthetic pathway as a vehicle for the display of cell-surface azides based on the efficient conversion of Ac₄ManNAz to SiaNAz observed in previous cell culture studies⁷. Furthermore, it has been reported that analogues of the natural substrate *N*-acetylmannosamine (ManNAc) with extended *N*-acyl chains are converted to the corresponding sialic acids in rats, suggesting that these precursors can access organs *in vivo*⁸. In cell culture, Ac₄ManNAz is a more efficient substrate for SiaNAz biosynthesis than free ManNAz owing to its facile entry into cells by passive diffusion¹⁷. The acetyl groups are then removed by carboxyesterases within the cell. Similar esterases exist at high levels in rodent serum^{18,19}, the activity of which might reduce the concentration of biologically available Ac₄ManNAz. Thus, we first investigated the conversion of Ac₄ManNAz to SiaNAz in the plasma esterase-deficient mouse strain Es1^e/Es1^e (refs 18, 20).

Es1^e/Es1^e mice were administered daily doses of Ac₄ManNAz (0–300 mg kg^{–1}) intraperitoneally for 7 days. On the eighth day, the mice were euthanized and their splenocytes (cells rich in sialosides) were isolated. The presence of cell-surface azides was quantified by Staudinger ligation *ex vivo* with a phosphine probe comprising a Flag peptide (Phos-Flag) (shown schematically in Fig. 1c)²¹. Treatment with a fluorescein isothiocyanate-labelled anti-Flag antibody

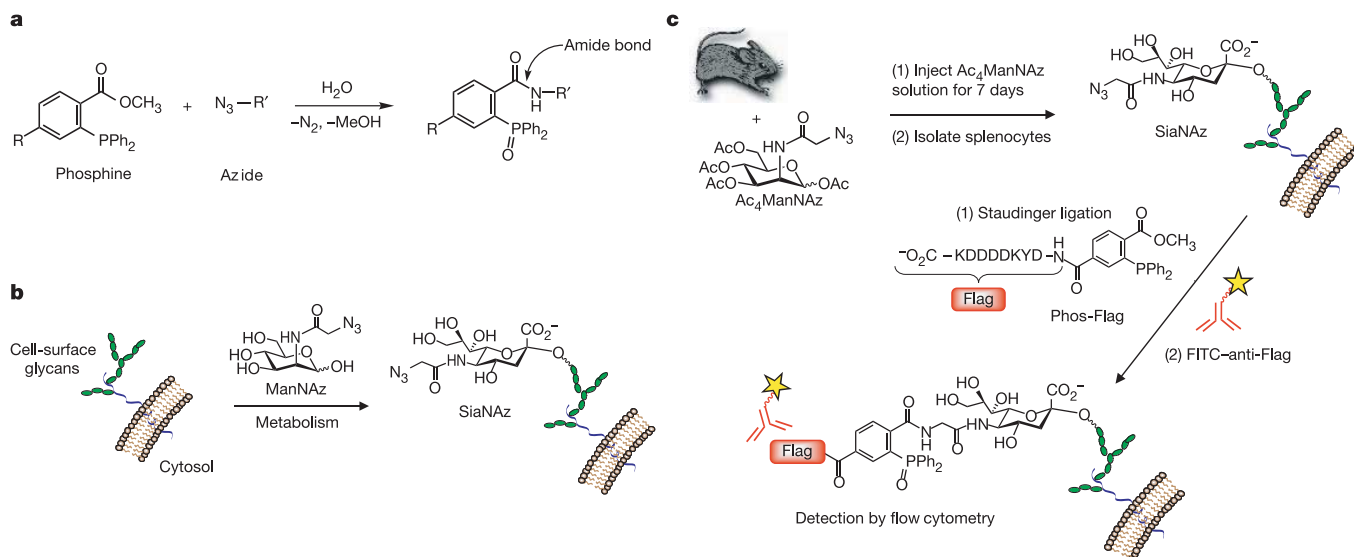


Figure 1 The Staudinger ligation and metabolic oligosaccharide engineering. **a**, The Staudinger ligation of an azide and functionalized phosphine results in the formation of an amide bond. **b**, Azides can be delivered to cell-surface glycoconjugates by metabolism of ManNAz to SiaNAz. **c**, Experimental overview for investigating the metabolic conversion of

Ac₄ManNAz *in vivo*. Splenocytes from mice treated with the azido sugar were collected and probed for the presence of cell-surface azides using Phos-Flag. Labelled cells were treated with FITC-anti-Flag and analysed by flow cytometry.

(FITC–anti-Flag) and subsequent flow cytometry analysis produced the data depicted in Fig. 2a, b. Splenocytes from mice treated with Ac₄ManNAz exhibited a dose-dependent increase in fluorescence, which was significantly higher than the signal observed with splenocytes from vehicle-treated mice. These data suggest that Ac₄ManNAz is metabolized to SiaNAz in murine splenocytes. Notably, treatment with the highest dose of Ac₄ManNAz produced no apparent adverse physiological effects over 7 days, as determined by monitoring feeding habits, weight and overall activity.

To assess the importance of peracetylation for the biological availability of ManNAz, we treated mice with Ac₄ManNAz or an equimolar amount of the free sugar. No azides were observed on splenocytes from mice treated with free ManNAz as assessed by flow cytometry, whereas Ac₄ManNAz produced a robust azide-dependent signal (Fig. 2c). Acetyl protection therefore considerably increases the efficiency of azido-sugar metabolism *in vivo*. Surprisingly, the extent of metabolic labelling in murine splenocytes was found to be identical in the plasma esterase-deficient Es1^e/Es1^e and wild-type B6D2F1 mice treated with Ac₄ManNAz (Fig. 2d). This finding suggests that a broad variety of experimental mouse models are amenable to chemical cell-surface remodelling.

We probed further for the presence of azido sugars in glycoproteins from a panel of murine organs. Selected organs were collected from mice that had been administered daily doses of Ac₄ManNAz (300 mg kg⁻¹) or vehicle for 7 days. These organs were homogenized, and the soluble fractions were reacted with Phos-Flag and analysed by western blot. As shown in Fig. 3a, labelled glycoproteins were only observed in organ lysates from mice exposed to Ac₄ManNAz. Discrete bands were observed from heart, kidney and liver lysates, but not from brain or thymus homogenates. The liver is known to secrete numerous sialylated glycoproteins and is a target of first-pass metabolism, factors that may explain the robust labelling observed in that organ. Notably, the kidney and heart lack significant levels of UDP-GlcNAc 2-epimerase²², the enzyme that produces endogenous ManNAc. Low levels of endogenous ManNAc might provide ManNAz with a

competitive advantage for metabolic labelling in these organs. Glycoprotein expression levels and tissue access might also contribute to the observed distribution of azides.

To confirm that SiaNAz is indeed the host of glycoprotein-associated azides, we incubated splenocytes from Ac₄ManNAz-treated and untreated mice with sialidase and evaluated the effects on cell-surface azide levels by Staudinger ligation with Phos-Flag. As shown in Fig. 3b, sialidase treatment significantly reduced the cell-surface fluorescence. Similarly, sialidase treatment of serum glycoproteins from mice administered Ac₄ManNAz reduced Phos-Flag labelling in a dose-dependent manner (Fig. 3c). Finally, direct characterization of SiaNAz was accomplished by sialic acid analysis of heart tissue lysates from Ac₄ManNAz-treated mice with com-

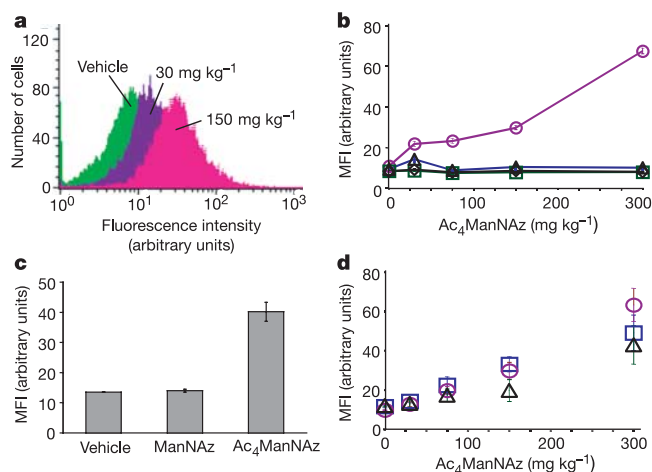


Figure 2 Ac₄ManNAz is metabolized *in vivo*. **a**, Flow cytometry analysis of splenocytes from Ac₄ManNAz-treated mice. **b**, Mean fluorescence intensity (MFI) of the cells from **a** as a function of azido-sugar dose (circles). Assay controls included unlabelled splenocytes from Ac₄ManNAz-treated mice (squares), splenocytes from Ac₄ManNAz-treated mice incubated with Phos-Flag followed by a class-matched control monoclonal antibody (diamonds), and splenocytes from Ac₄ManNAz-treated mice incubated with FITC–anti-Flag only (triangles). **c**, MFI of splenocytes from Ac₄ManNAz- and ManNAz-treated Es1^e/Es1^e mice. **d**, MFI of splenocytes from Es1^e/Es1^e mice (triangles) or wild-type B6D2F1 mice (circles, males; squares, females) treated with Ac₄ManNAz. Error bars represent the standard deviation of the mean for three replicate Staudinger ligation reactions. For **a–d** similar results were obtained in two replicate experiments.

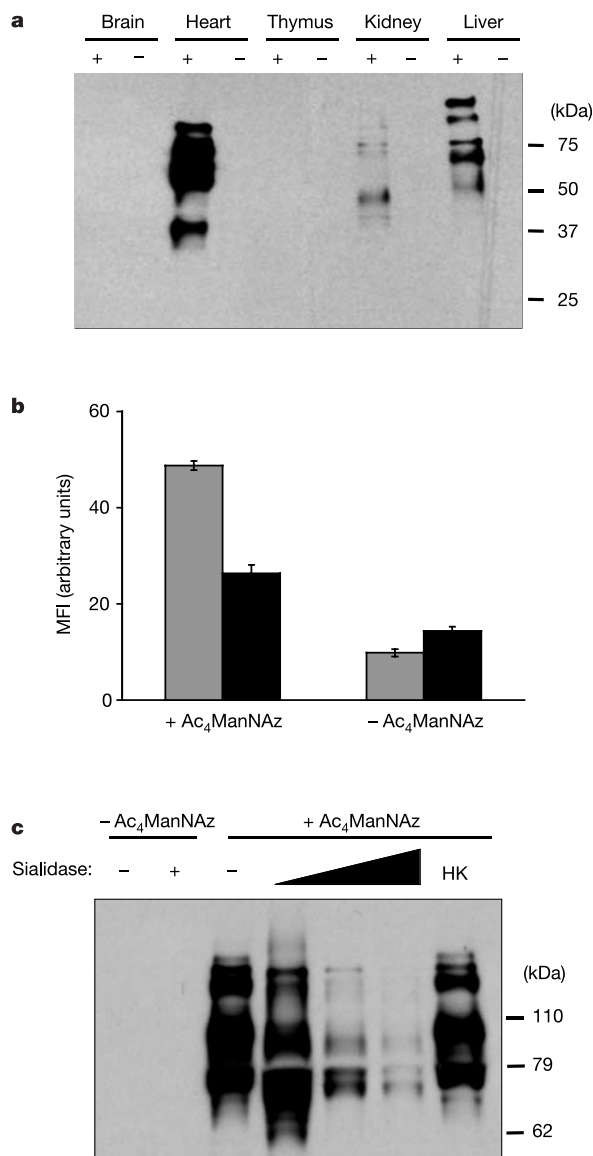


Figure 3 Analysis of SiaNAz on cells and in tissues. **a**, Western blot analysis of tissue lysates from Es1^e/Es1^e mice administered Ac₄ManNAz (+) or vehicle alone (-). The same patterns of labelling were apparent in several experiments. **b**, Splenocytes from B6D2F1 mice treated with Ac₄ManNAz or vehicle were treated with *A. ureafaciens* sialidase (black bars) or left untreated (grey bars). The cells were analysed by flow cytometry; error bars represent the standard deviation of the mean for three replicate Staudinger ligation reactions. **c**, Western blot analysis of sialidase-treated (+) or untreated (-) serum samples from mice administered Ac₄ManNAz or vehicle. The samples were incubated with active or heat-killed (HK) sialidase. Total protein loading was confirmed by Coomassie-stained protein gel (not shown).

parison to an authentic sample of SiaNAz²³. Quantification of the total sialic acid population revealed that SiaNAz had replaced approximately 3% of natural sialic acids in this tissue (data not shown). This value is sufficient for detection but unlikely to grossly perturb sialic-acid-dependent interactions. Importantly, no major differences were observed in the sialoglycoprotein profiles of organs from Ac₄ManNAz-treated and untreated mice (Supplementary Fig. 1).

Cells coated with SiaNAz *in vivo* are primed for further chemical remodelling within their native physiological environment. To achieve this in practice would require that the Staudinger ligation function on cells within living animals, an unprecedented event for a synthetic reaction. The demands on any reaction in this context are considerable. Aside from displaying extraordinary chemical selectivity, the reactants must possess exquisite biological compatibility; that is, they must not produce harmful side effects nor be prone to rapid metabolic breakdown on the timescale of the reaction. The Staudinger ligation is uniquely poised to fulfil these criteria because its two components, the azide and the phosphine, have been used previously *in vivo*. For example, the azide is a component of the well-known drug AZT (azidothymidine), and phosphine-metal conjugates are established therapeutic and diagnostic agents^{24,25}.

To test the Staudinger ligation in living animals, we administered Ac₄ManNAz (300 mg kg⁻¹) intraperitoneally to C57BL/6 mice once daily for 7 days. On the eighth day, mice were injected intraperitoneally with 16 μmol (~1 molar equivalent on the basis of a single dose of 300 mg kg⁻¹ Ac₄ManNAz) of Phos-Flag. After 90 min, the mice were euthanized, splenocytes were isolated, and cell-surface Flag epitopes were probed by flow cytometry. Only splenocytes from mice treated with both Ac₄ManNAz and Phos-Flag displayed a significant increase in fluorescence relative to splenocytes from untreated mice, indicating that the Staudinger ligation had proceeded *in vivo* (Fig. 4, grey bars). We performed similar experiments in which mice were exposed to Phos-Flag for longer time periods before splenocyte analysis. The labelling intensity was significantly reduced after 12 h and no significant labelling was detected after 24 h (data not shown). This time-dependent reduction in labelling might be attributed to degradation of the Flag peptide or to turnover of membrane-associated glycans. No adverse physiological effects were observed even after 24 h of exposure to Phos-Flag, as judged by feeding behaviour and overall activity.

To determine the extent of the *in vivo* Staudinger ligation after 90 min, we subjected harvested splenocytes to further reaction with Phos-Flag *ex vivo* and analysed them by flow cytometry. As shown in Fig. 4 (black bars), cells from mice treated with Ac₄ManNAz

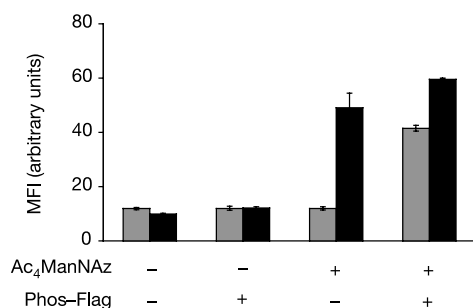


Figure 4 The Staudinger ligation proceeds *in vivo*. Mice were administered Ac₄ManNAz or vehicle once daily for 7 days. On the eighth day, the mice were administered Phos-Flag (16 μmol in ~200 μl PBS) or an equal volume of vehicle. After 1.5 h, splenocytes were treated with FITC-anti-Flag and analysed by flow cytometry (grey bars). A portion of the isolated splenocytes was further reacted with Phos-Flag and analysed as in Fig. 2 (black bars). Error bars represent the standard deviation of the mean for three replicate FITC-anti-Flag labelling reactions or Staudinger ligation reactions.

alone displayed an increase in fluorescence commensurate with the presence of unligated azides. Cells that had been treated with both Ac₄ManNAz and Phos-Flag showed a less pronounced increase in fluorescence upon *ex vivo* ligation, suggesting that a significant proportion of the available azides had been ligated *in vivo*. The products of the Staudinger ligation were also observed on serum glycoproteins from mice treated with Ac₄ManNAz and Phos-Flag by western blot analysis (Supplementary Fig. 2).

The ability to chemically modify cell-surface glycans in living animals provides a means to monitor these biopolymers in a physiologically relevant system. For example, the Staudinger ligation might be used to target probes for non-invasive imaging to cells as a function of their glycosylation pattern. Changes in glycosylation associated with stages of organ development or disease progression could be visualized in this fashion. Notably, elevated levels of sialylated glycans have been observed on numerous cancers^{15,26,27} and at sites of inflammation¹⁶. More broadly, the azide can serve as an *in vivo* reporter of secondary metabolite expression. Lipids, steroids and cofactors could potentially be probed if their metabolic enzymes are tolerant of azido precursors.

The range of opportunities provided by the Staudinger ligation underscores the potential impact of bio-orthogonal reactions that can be carried out in whole organisms. An ongoing challenge in the chemistry community is to identify new transformations with the requisite qualities of selectivity and biocompatibility. Indeed, the azide has an alternative mode of reactivity, the 1,3-dipolar cycloaddition with alkynes, which has been used for bioconjugation reactions and could potentially be modified for use *in vivo*²⁸. The Staudinger ligation may therefore be the first in a future arsenal of chemical reactions used to probe biology in living animals. □

Methods

Compound administration

Ac₄ManNAz, ManNAz and Phos-Flag were synthesized according to previously published procedures^{7,21}. For metabolic labelling experiments, mice (Es1^{+/+}Es1⁻, B6D2F1 or C57BL/6) were administered daily doses of Ac₄ManNAz (0–300 mg kg⁻¹ in ~200 μl of 70% aqueous DMSO, from a stock solution of 50 mg ml⁻¹) or ManNAz (0–182 mg kg⁻¹ in H₂O) intraperitoneally for 7 days. Organs were collected 24 h after the final azido-sugar injection. To test the Staudinger ligation in living animals, C57BL/6 mice were administered Ac₄ManNAz (300 mg kg⁻¹) or vehicle (70% DMSO) intraperitoneally once daily for 7 days. Twenty-four hours after the final Ac₄ManNAz bolus, mice were injected intraperitoneally with Phos-Flag (16 μmol in ~200 μl PBS) or vehicle (PBS). Organs were collected 90 min after the Phos-Flag injection.

Labelling of splenocyte cell-surface azides *ex vivo*

After mice were administered Ac₄ManNAz or vehicle alone, their splenocytes were isolated using a standard protocol and probed for the presence of cell-surface azides using a Staudinger ligation assay⁹. Briefly, splenocytes were incubated with 250 μM Phos-Flag for 1 h at room temperature, then treated with FITC-anti-Flag or FITC-conjugated mouse IgG₁ isotype control for 30 min on ice and analysed by flow cytometry.

Lysis of murine organs and western blot analysis

Isolated murine organs were rinsed with PBS (pH 7.4) and homogenized in 2 ml of lysis buffer²⁹. To probe for the presence of azides, aliquots of the tissue lysates were diluted 1:1 with 500 μM Phos-Flag and incubated at room temperature for 6–12 h. The samples were analysed by western blot probing with horseradish peroxidase (HRP)-anti-Flag¹¹.

Sialidase treatment of splenocytes and organ lysates

Splenocytes from Ac₄ManNAz-treated (300 mg kg⁻¹ intraperitoneally once daily for 7 days) or untreated B6D2F1 mice were incubated with *Arthrobacter ureafaciens* sialidase (Roche, 20 mU, 100 μl final volume) in sialidase buffer³⁰. After 1 h at room temperature, the cells were treated with Phos-Flag followed by FITC-anti-Flag as described above. After a 30 min incubation on ice, the cells were washed and analysed by flow cytometry.

For western blot analysis, murine serum samples were combined with buffer (154 mM NaCl, 50 mM sodium acetate, 9 mM CaCl₂, pH 5.5) and *Vibrio cholerae* sialidase (Calbiochem, 0–37 mU, dissolved in the same buffer system) to a total volume of 20 μl. The samples were incubated at 37 °C overnight, then diluted 1:1 with 500 μM Phos-Flag (final Phos-Flag concentration = 250 μM) and incubated at room temperature for 8 h. The samples were further analysed by western blot as described above.

Splenocyte analysis after *in vivo* Staudinger ligation

Splenocytes from C57BL/6 mice treated with Ac₄ManNAz or vehicle (70% DMSO) and Phos-Flag or vehicle (PBS) were isolated and probed for the presence of cell-surface Flag epitopes. Briefly, splenocytes were incubated directly with FITC-anti-Flag (1:900 dilution)

for 30 min on ice. Alternatively, the splenocytes were treated with Phos-Flag *ex vivo* and then FITC-anti-Flag as described above. All cells were analysed by flow cytometry.

Received 12 May; accepted 25 June 2004; doi:10.1038/nature02791.

1. Keppler, O. T., Horstkorte, R., Pawlita, M., Schmitts, C. & Reutter, W. Biochemical engineering of the *N*-acyl side chain of sialic acid: biological implications. *Glycobiology* **11**, 11R–18R (2001).
2. Dube, D. H. & Bertozzi, C. R. Metabolic oligosaccharide engineering as a tool for glycobiology. *Curr. Opin. Chem. Biol.* **7**, 616–625 (2003).
3. Buttner, B. *et al.* Biochemical engineering of cell surface sialic acids stimulates axonal growth. *J. Neurosci.* **22**, 8869–8875 (2002).
4. Keppler, O. T. *et al.* Biosynthetic modulation of sialic acid-dependent virus-receptor interactions of two primate polyoma viruses. *J. Biol. Chem.* **270**, 1308–1314 (1995).
5. Charter, N. W., Mahal, L. K., Koshland, D. E. & Bertozzi, C. R. Differential effects of unnatural sialic acids on the polysialylation of the neural cell adhesion molecule and neuronal behavior. *J. Biol. Chem.* **277**, 9255–9261 (2002).
6. Mahal, L. K., Yarema, K. J. & Bertozzi, C. R. Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis. *Science* **276**, 1125–1128 (1997).
7. Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **287**, 2007–2010 (2000).
8. Kayser, H. *et al.* Biosynthesis of a nonphysiological sialic acid in different rat organs, using *N*-propanoyl-D-hexosamines as precursors. *J. Biol. Chem.* **267**, 16934–16938 (1992).
9. Saxon, E. *et al.* Investigating cellular metabolism of synthetic azidosugars with the Staudinger ligation. *J. Am. Chem. Soc.* **124**, 14893–14902 (2002).
10. Hang, H. C., Yu, C., Kato, D. L. & Bertozzi, C. R. A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl Acad. Sci. USA* **100**, 14846–14851 (2003).
11. Vocado, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl Acad. Sci. USA* **100**, 9116–9121 (2003).
12. Kohn, M. & Breinbauer, R. The Staudinger ligation—A gift to chemical biology. *Angew. Chem. Int. Edn Engl.* **43**, 3106–3116 (2004).
13. Ova, H. *et al.* Chemistry in living cells: Detection of active proteasomes by a two-step labeling strategy. *Angew. Chem. Int. Edn Engl.* **42**, 3626–3629 (2003).
14. Kim, Y. J. & Varki, A. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconj. J.* **14**, 569–576 (1997).
15. Fukuda, M. Possible roles of tumor-associated carbohydrate antigens. *Cancer Res.* **56**, 2237–2244 (1996).
16. Renkonen, J., Tynnenen, O., Hayry, P., Paavonen, T. & Renkonen, R. Glycosylation might provide endothelial zip codes for organ-specific leukocyte traffic into inflammatory sites. *Am. J. Pathol.* **161**, 543–550 (2002).
17. Luchansky, S. J. *et al.* Constructing azide-labeled cell surfaces using polysaccharide biosynthetic pathways. *Methods Enzymol.* **362**, 249–272 (2003).
18. Kavarana, M. J., Kovaleva, E. G., Creighton, D. J., Wollman, M. B. & Eisman, J. L. Mechanism-based competitive inhibitors of glyoxalase I: Intracellular delivery, *in vitro* antitumor activities, and stabilities in human serum and mouse serum. *J. Med. Chem.* **42**, 221–228 (1999).
19. Morton, C. L. *et al.* Activation of CPT-11 in mice: Identification and analysis of a highly effective plasma esterase. *Cancer Res.* **60**, 4206–4210 (2000).
20. Soares, E. R. Identification of a new allele of Es-I segregating in an inbred strain of mice. *Biochem. Genet.* **17**, 577–583 (1979).
21. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl Acad. Sci. USA* **99**, 19–24 (2002).
22. Stasche, R. *et al.* A bifunctional enzyme catalyzes the first two steps in *N*-acetylneuraminic acid biosynthesis of rat liver—Molecular cloning and functional expression of UDP-*N*-acetyl-glucosamine 2-epimerase/*N*-acetylmannosamine kinase. *J. Biol. Chem.* **272**, 24319–24324 (1997).
23. Luchansky, S. J., Argade, S., Hayes, B. K. & Bertozzi, C. R. Metabolic functionalization of recombinant glycoproteins. *Biochemistry* (in the press).
24. Shaw, C. F. Gold-based therapeutic agents. *Chem. Rev.* **99**, 2589–2600 (1999).
25. Jurisson, S. S. & Lydon, J. D. Potential technetium small molecule radiopharmaceuticals. *Chem. Rev.* **99**, 2205–2218 (1999).
26. Codrington, J. F., Klein, G., Silber, C., Linsley, K. B. & Jeanloz, R. W. Variations in the sialic acid compositions in glycoproteins of mouse ascites tumor-cell surfaces. *Biochemistry* **18**, 2145–2149 (1979).
27. Sell, S. Cancer-associated carbohydrates identified by monoclonal antibodies. *Hum. Pathol.* **21**, 1003–1019 (1990).
28. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–1137 (2003).
29. Martin, D. C., Fowlkes, J. L., Babic, B. & Khokha, R. Insulin-like growth factor II signaling in neoplastic proliferation is blocked by transgenic expression of the metalloproteinase inhibitor TIMP-1. *J. Cell Biol.* **146**, 881–892 (1999).
30. Reicher, J. S., Whiteheart, S. W. & Hart, G. W. Intracellular trafficking of cell surface sialoglycoconjugates. *J. Biol. Chem.* **263**, 16316–16326 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Jamieson, S. Luchansky, H. Hang and P. Drake for discussions. J.A.P. was supported by a HHMI Predoctoral Fellowship and D.H.D. was supported by a National Science Foundation Predoctoral Fellowship. Sialic acid analysis was performed by the UCSD GRTC Glycotechnology Core Resource. This work was supported by grants from Johnson & Johnson (Focused Giving Grant), the Mizutani Foundation for Glycoscience, the US Department of Energy and the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.B. (crb@berkeley.edu).

Variable ageing and storage of dissolved organic components in the open ocean

Ai Ning Loh^{1*}, James E. Bauer¹ & Ellen R. M. Druffel²

¹*School of Marine Science, College of William and Mary, PO Box 1346, Gloucester Point, Virginia 23062, USA*

²*Department of Earth System Science, University of California at Irvine, 3224 Croul Hall, Irvine, California 92697-3100, USA*

* Present address: Division of Ecological Studies, College of Arts and Sciences, Florida Gulf Coast University, 10501 FGCU Boulevard South, Fort Myers, Florida 33965, USA

Seawater dissolved organic matter (DOM) is the largest reservoir of exchangeable organic carbon in the ocean, comparable in quantity to atmospheric carbon dioxide^{1,2}. The composition, turnover times and fate of all but a few planktonic constituents of this material are, however, largely unknown^{3,4}. Models of ocean carbon cycling are thus limited by the need for information on temporal scales of carbon storage in DOM subcomponents, produced via the 'biological pump' relative to their recycling by bacteria^{3,4}. Here we show that carbohydrate- and protein-like substances in the open Atlantic and Pacific oceans, though often significantly aged, comprise younger fractions of the DOM, whereas dissolved lipophilic material exhibits up to ~90 per cent fossil character. In contrast to the millennial mean ages of DOM observed throughout the water column, weighted mean turnover times of DOM in the surface ocean are only decadal in magnitude. An observed size-age continuum further demonstrates that small dissolved molecules are the most highly aged forms of organic matter, cycling much more slowly than larger, younger dissolved and particulate precursors, and directly links oceanic organic matter age and size with reactivity^{3,5}.

Seawater DOM consists of analytically identifiable biochemicals such as carbohydrates, proteins and lipids, as well as operationally defined and long-lived geomacromolecules (for example, humic and fulvic substances^{5,6}). In order to resolve some of the key details of DOM sources and cycling in the oceans, major organic components were extracted from high-molecular-weight ultrafiltered DOM⁵ (DOM_{HMW} > 1,000 daltons) collected from 1,000–3,000 l of sea water, and analysed for both $\Delta^{14}\text{C}$ and $\delta^{13}\text{C}$ isotopic signatures. Samples were collected from surface mixed-layer (3–20 m), mesopelagic oxygen-minimum (850–900 m), and abyssal (1,500–1,800 m) depths in the central North Pacific (June 1999) and the Sargasso Sea region of the North Atlantic (June 2000) oligotrophic ocean gyres. The contributions of solvent-extractable lipids, protein-like and carbohydrate-like organic matter (OM), as well as different molecular-weight fractions, to the overall age structure of seawater DOM, were thus established.

By far the most highly aged DOM component was the lipid extract (6.4–17.1 kyr before present, BP; Table 1), with ¹⁴C ages in the deep Pacific representing the greatest yet observed for any component of seawater OM. The lipid extract was considerably older by ~5–13 kyr than the total DOM_{HMW} and unfractionated, bulk DOM (ΣDOM) pools (Tables 1 and 2; Fig. 1a). Furthermore, at all mesopelagic and abyssal depths, the lipid extract and DOM_{HMW} were older in the Pacific than in the Atlantic, similar to the ocean-ocean offsets observed for ΣDOM ^{6,7} (Table 1) and presumably due to cumulative ageing during deep water-mass transit⁸. Conversely, mixed-layer lipid extract, DOM_{HMW} and ΣDOM were all older in the Atlantic than in the Pacific (Tables 1 and 2), suggesting possible aged North American continental or atmospheric inputs there⁹. The highly $\delta^{13}\text{C}$ -depleted signatures of lipid extracts (Table 1, Fig. 1b) are consistent with isotopic fractionations during cellular lipid

synthesis¹⁰. Thus, these fossil dissolved lipids arise either from extensive ageing within the oceans, with concomitant recycling over many ocean circulation times⁸, or from inputs of one or more pre-aged lipophilic precursors¹¹. Concurrent elemental ratios (Table 2) and lipid biomarker data¹² of total DOM_{HMW} suggest, however, that the lipid extract may be dominated by planktonic, rather than petrogenic, material.

Dissolved protein-like and carbohydrate-like fractions were similar in ¹⁴C age, ranging from modern to ~3–4 kyr BP (Table 1). Modern to near-modern ages in surface waters indicate that both fractions are derived from recent, post-bomb (that is, after ~1955) marine production, and contain little aged or recycled material. In deeper waters, however, these components, which are quite reactive in surface waters⁴, are deduced to have escaped degradation over many ocean circulation times⁸ and to have aged extensively. The protein- and carbohydrate-like components were younger by as much as ~13–14 kyr compared to the corresponding lipid extract (Table 1), and by as much as ~1 kyr (Fig. 1a) compared to DOM_{HMW}, supporting the contention that 'old' seawater ΣDOM^{1,6,7} is actually composed of components having a spectrum of ages and reactivities. Transport-based ageing of protein- and

carbohydrate-like DOM is also suggested at mesopelagic and abyssal depths (Table 1), similar to the ΣDOM^{6,7}, total DOM_{HMW} and lipid extract (Tables 1 and 2).

Dissolved forms of all three major organic fractions were significantly older than their particulate counterparts^{13,14} (Fig. 2). Dissolved lipid extracts were ~6–17 kyr older than lipid extracts of sinking particulate OM (POM; 3,500 m depth)^{13,14}, suggesting that dissolved and particulate lipids cycle on dramatically different timescales or arise from dissimilar sources. The latter possibility is supported by correspondingly lower δ¹³C of dissolved lipids (about –29 to –28‰; Table 1) compared to particulate and sedimentary lipids (–25‰ to –22‰; refs 13, 14; Fig. 2). Alternatively, particulate lipids may contain both a highly aged component, similar to the dissolved pool, and a recently derived component from contemporary marine production¹⁵, accounting for their intermediate ages. Deep dissolved protein- and carbohydrate-like fractions were also significantly older by ~3–4 kyr than particulate forms^{13,14}, whereas the relative abundances of these two fractions are reversed in the dissolved (Table 1) compared to particulate phases^{13,14}, suggesting dissimilar sources or preservation. Therefore, although the dissolved fractions are far more abundant, they are also far longer-lived

Table 1 Isotopic signatures of organic fractions contained in seawater DOM_{HMW}

Depth	LE			PL			CL			MUC		
	%*	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	%*	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	%*	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	%†	Δ ¹⁴ C (‰)	δ ¹³ C (‰)
Atlantic												
3 m	0.1	–637 (8,140)§	–28‡	13.3	2 (modern)	–21.2	47.5	13 (modern)	–21.5	39.1	–28 (230)	–22.3
850 m	0.3	–730 (10,520)	–28.1	17.1	–190 (1,690)	–20.4	38.9	–228 (2,080)	–20.4	43.7	–336 (3,292)	–21.7
1,500 m	0.2	–830 (14,250)	–28‡	22.9	–215 (1,950)	–20.8	33.8	–309 (2,969)	–21‡	43.0	–246 (2,271)	–21.5
Pacific												
20 m	0.3	–551 (6,430)	–27.6	19.0	–21 (170)	–21‡	35.9	7 (modern)	–21.4	44.9	–199 (1,780)	–22.3
900 m	0.3	–864 (16,010)	–29.4	15.4	–279 (2,620)	–21.0	27.7	–302 (2,890)	–20.4	56.7	–444 (4,720)	–22.1
1,800 m	0.3	–881 (17,100)	–28‡	16.9	–332 (3,240)	–20.8	41.1	–406 (4,180)	–20.3	41.7	–499 (5,550)	–22.3

*% of each organic fraction in DOM_{HMW}, based on C equivalents measured after combustion to CO₂.

†%MUC = 100 – (%LE + %PL + %CL). Isotopic signatures of MUC were estimated by isotopic mass balance, as: $X_{MUC} = (X_{HMW} - f_{LE}X_{LE} - f_{PL}X_{PL} - f_{CL}X_{CL})/f_{MUC}$, where X is the Δ¹⁴C or δ¹³C signature of each organic fraction, f is the relative contribution of each to the total DOM_{HMW} pool, and $f_{LE} + f_{PL} + f_{CL} + f_{MUC} = 1.0$.

‡Assumed from average of all observed values in that group for purposes of inclusion of Δ¹⁴C–δ¹³C pair in Fig. 2.

§Values in parentheses are ¹⁴C ages in yr BP calculated from Δ¹⁴C values as: Age (yr BP) = –8,033 ln(1 + Δ¹⁴C/_{1,000}) where yr BP indicate years before 1950, prior to thermonuclear weapons testing²⁷. As Δ¹⁴C values were not reservoir-corrected, all ages are not true calendar ages.

Table 2 Isotopic and elemental characteristics of seawater DOM size classes

Depth	ΣDOM*		DOM _{HMW} †		DOM _{LMW} ‡		DOM _{HMW}		
	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	Δ ¹⁴ C (‰)	δ ¹³ C (‰)	C:N	C:P	N:P
Atlantic									
3 m	–238 (2,180)§	–21.3	–5 (40)	–21.8	–280 (2,640)	–21.2	14	353	25
850 m	–375 (3,780)	–21.2	–270 (2,528)	–21	–383 (3,885)	–21.2	14	268	19
1,500 m	–378 (3,810)	–20.8	–262 (2,440)	–21.2	–384 (3,890)	–20.8	14	196	14
Pacific									
20 m	–191 (1,700)	–21.2	–92 (780)	–21.8	–210 (1,890)	–21.1	15	280	19
900 m	–470 (5,100)	–20.8	–381 (3,850)	–21.5	–481 (5,260)	–20.7	14	284	21
1,800 m	–533 (6,120)	–21.2	–434 (4,570)	–21.3	–539 (6,220)	–21.2	14	256	18

*Unfractionated, bulk DOM values^{6,7}.

†High-molecular-weight DOM was 15–16% of ΣDOM in surface waters, and 5–11% in deeper waters based on C recoveries after ultrafiltration, diafiltration and lyophilisation. These recoveries are consistent with the range of recoveries reported in other studies^{5,16,19}.

‡Isotopic signatures of low-molecular-weight DOM, estimated by isotopic mass balance as: $X_{LMW} = (X_{ΣDOM} - f_{HMW}X_{HMW})/(1 - f_{HMW})$, where X is the Δ¹⁴C or δ¹³C value of each molecular weight fraction, f is the relative contribution of HMW and LMW material to ΣDOM, and $f_{HMW} + f_{LMW} = 1.0$.

§Values in parentheses are ¹⁴C ages in yr BP calculated from Δ¹⁴C values as in Table 1.

||Assumed from average of all observed values in that group for purposes of inclusion of Δ¹⁴C–δ¹³C pair in Fig. 2.

and are thus deduced to be less reactive to bacteria than the younger particulate fractions. Dissolved carbohydrates have further been found to contain specific sugars of modern ^{14}C age¹⁶, indicating that even within a given organic fraction individual molecules may have unique cycling times, similar to findings for individual sedimentary lipids¹¹.

The highly modified, acid-insoluble fraction of DOM_{HMW} (analogous to the molecularly uncharacterized, MUC, component of OM^{17}) was estimated to comprise 39–57% of DOM_{HMW} carbon (Table 1). These large amounts of MUC-like DOM and its estimated ages (0.2–5.6 kyr BP; Table 1) suggest that much of the ΣDOM is composed of structurally modified biopolymers and geomolecules¹⁷, probably derived from combinations of diagenetically altered younger proteins and sugars and older lipid components. Based on their disparate $\Delta^{14}\text{C}$ and $\delta^{13}\text{C}$ signatures (Table 1, Fig. 2), however, the dissolved forms of MUC and lipid extract are unlikely to share a common origin as has been suggested for POM^{14} . Instead, dissolved MUC in deep waters is isotopically more similar to, and thus may arise from, dissolved protein- and carbohydrate-like and humic materials⁶ (Fig. 2), or from a precursor common to each. This is further supported by NMR studies demonstrating that DOM_{HMW} contains significant acyl polysaccharide, a carbohydrate-derived biopolymer¹⁸.

Radiocarbon ages for total DOM_{HMW} ranged widely, from modern to 4.6 kyr BP (Table 2), and were younger than $\Sigma\text{DOM}^{1,6}$ from the Atlantic and Pacific (Table 2). Therefore, ΣDOM must by definition also contain an older, low-molecular-weight (LMW) component in order to balance the younger DOM_{HMW} (ref. 19; Table 2). In addition to being the oldest size fraction yet identified for seawater OM (~ 1.9 –6.2 kyr BP), DOM_{LMW} is also the most

abundant, comprising 77–95% of ΣDOM . The $\delta^{13}\text{C}$ signatures of this older LMW material further suggest that it arises directly from recycling of younger DOM_{HMW} (Table 2). Observations of old DOM_{LMW} , intermediate-aged DOM_{HMW} , and young $\text{POM}^{7,20}$ in the oceans reveal a pronounced size–age relationship among the major forms of seawater OM (Table 2; Fig. 2). The two main forms of POM (that is, sinking and suspended) consistently contain bomb ^{14}C ($\Delta^{14}\text{C}$ range: about -100‰ to $+160\text{‰}$ for suspended POM, and about -30‰ to $+35\text{‰}$ for sinking POM) throughout the Pacific and Atlantic water columns^{7,20}. However, only surface DOM_{HMW} is similarly enriched, while LMW material contains no apparent bomb ^{14}C at any depth (Table 2; Fig. 2). That is, ^{14}C age increases consistently in the size sequence from sinking POM (the youngest, largest fraction), to suspended POM, to DOM_{HMW} and finally DOM_{LMW} (the oldest, smallest fraction; Fig. 2), and most probably arises from sequential hydrolysis, dissolution and/or degradation of larger forms of OM to successively smaller fractions. This also suggests that the relative rate of OM respiration slows as highly aged LMW material accumulates in the latter stages of degradation.

An important corollary of this size–age continuum is that it coincides with a previously observed OM size–reactivity continuum^{3,5}, ranging from structurally complex and recently produced sinking POM and DOM_{HMW} (most reactive to bacterial degradation) to structurally simple but highly reworked DOM_{LMW} (least reactive to bacterial degradation). The proposed size–age model for seawater OM is therefore consistent with higher utilization rates of DOM_{HMW} than DOM_{LMW} in oceanic and coastal waters^{3,5} as well as with the presence of younger, presumably more reactive sub-components^{3,16} in DOM_{HMW} (Fig. 1a). Studies of ΣDOM degradation further support the presence of specific bioavailable sub-fractions of this large, heterogeneous pool⁴. Besides the ^{14}C evidence, highly elevated elemental ratios of DOM_{HMW} (Table 2; ref. 21) compared to recently produced ‘Redfield’ OM (C:N:P $\sim 106:16:1$), together with the presence of bacterial fatty acids in oceanic DOM_{HMW} (ref. 12), further support the contention that dissolved components may have undergone extensive recycling compared to POM (Fig. 2). An alternative to the proposed size-dependent ageing model for seawater OM is that there exist one or

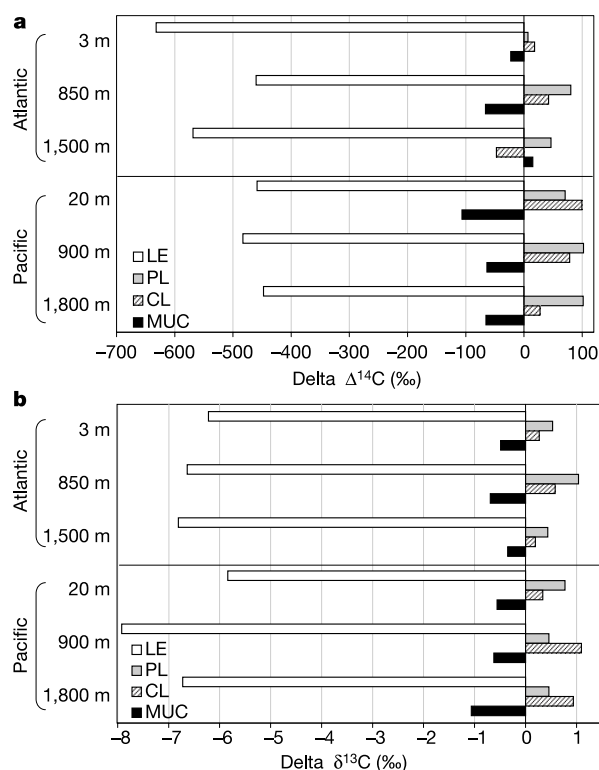


Figure 1 Isotopic signatures of dissolved organic fractions relative to DOM_{HMW} . **a**, Plotted for $\Delta^{14}\text{C}$ (as $\text{Delta } \Delta^{14}\text{C}$); and **b**, plotted for $\Delta^{13}\text{C}$ (as $\text{Delta } \delta^{13}\text{C}$) for three depths each in the Sargasso Sea region of the North Atlantic and in the central North Pacific. LE, lipid-extract fraction; PL, protein-like fraction; CL, carbohydrate-like fraction; MUC, molecularly uncharacterized fraction. See Table 1 for assumed $\delta^{13}\text{C}$ values.

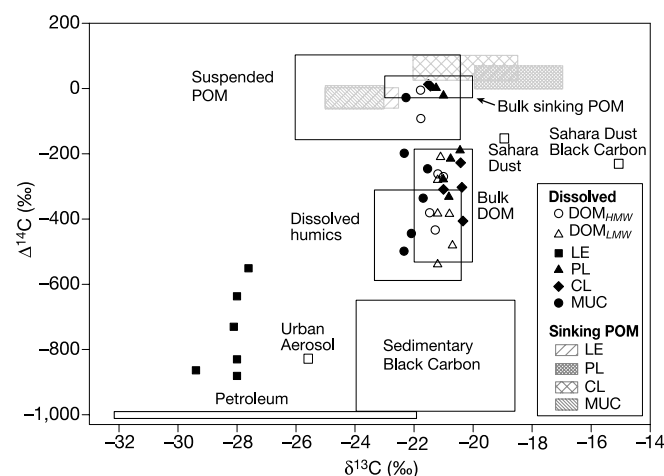


Figure 2 $\Delta^{14}\text{C}$ versus $\delta^{13}\text{C}$ for dissolved organic fractions, DOM_{HMW} , and DOM_{LMW} . Data from the present study are plotted as discrete points, along with ranges for potential sources of DOM in the North Atlantic and North Pacific oceans (boxes) compiled from the literature as follows: sinking POM⁷; suspended POM²⁰; bulk $\text{DOM}^{6,7}$; dissolved humic substances⁶; Sahara dust bulk organic matter and black carbon fraction⁹; urban aerosol²⁹; North Pacific and Southern Ocean sedimentary black carbon²²; and petroleum³⁰. Shaded boxes show $\Delta^{14}\text{C}$ and $\delta^{13}\text{C}$ ranges for organic fractions isolated from sinking POM^{13,14}. Abbreviations as in Fig. 1. See Tables 1 and 2 for assumed $\delta^{13}\text{C}$ values.

more allochthonous or autochthonous sources of pre-aged DOM in the oceans that are largely independent of plankton-dominated OM formation and degradation. While aged allochthonous sources of DOM such as submicrometre fossil black carbon²² are probably minimal on the basis of the observed OM size-age distributions (Fig. 2), other 'pre-aged' inputs may include natural hydrocarbon seepage²³ in certain ocean regions and atmospheric deposition⁹. Soluble forms of seawater OM are further predicted to escape degradation and undergo ageing either within the deep ocean, or during surface-to-deep ocean transport, by mechanisms different from those such as sorptive preservation²⁴ identified for particulate and sedimentary forms.

Within the oceanic DOM pool, various organic and size fractions persist on radically different timescales (Tables 1 and 2). The turnover time (TOT) of undifferentiated bulk pools such as Σ DOM is equivalent to its ^{14}C age if substances in the bulk pool are uniform in age. However, the presence of discrete organic and size fractions having different ^{14}C ages (and therefore different rates of turnover) results in a weighted mean TOT that departs from the Σ DOM age²⁵. On the basis of the heterogeneity in the organic and size fraction ages (Tables 1 and 2), the weighted mean TOT for surface ocean Σ DOM is estimated to be ~60–90 yr, increasing to ~3,700–6,000 yr in Atlantic and Pacific deep waters, respectively (see Supplementary Table S1). Surface ocean differences between the weighted mean TOT (decadal) and Σ DOM ^{14}C ages (millennial) result from surface DOM being dominated by young protein- and carbohydrate-like fractions that are recycled rapidly compared to the balance of the DOM. This contrasts with deep waters, where Σ DOM ages and TOT converge owing to the uniform low reactivity and turnover of all subcomponents. □

Methods

Sample collection and organic fraction separation

Large-volume (1,000–3,000 l) water samples were collected using 30-l rosette-mounted Niskin bottles. Samples were pre-filtered (0.2- μm) and transferred to an Amicon DC-10L tangential flow ultrafiltration system equipped with spiral wound filter cartridges with a 1,000 dalton molecular-weight cut-off. Samples were reduced to ~1 l and frozen until processing. For analyses, samples were thawed, diafiltered to remove salts, and lyophilised, followed by sequential extraction for solvent-extractable lipids, protein-like and carbohydrate-like organic fractions. Total lipids were extracted from the lyophilised sample using a modified Bligh-Dyer extraction with dichloromethane:methanol (2:1 v/v) and a Dionex accelerated solvent extractor¹². Residue from the lipid extraction was divided by weight into two portions for protein-like and carbohydrate-like extraction and isotopic analyses (adapted from ref. 13). One portion was hydrolysed (6 N HCl at 100 °C for 19 h) and eluted with 1.5 N NH_4OH through a cation exchange column to trap the protein-like fraction. The other portion was hydrolysed for the carbohydrate-like fraction (72% H_2SO_4 for 2 h, then 0.6 M H_2SO_4 at 100 °C for 2 h). The solution was neutralized with $\text{Ba}(\text{OH})_2 \cdot 8\text{H}_2\text{O}$, adjusted to pH 6–7 with 1.5 N NH_4OH , and eluted with Nanopure water through a cation/anion exchange column to trap the carbohydrate-like fraction.

Concentrations and carbon isotope signatures of known standard compounds were measured before and after the extraction procedure to evaluate extraction efficiencies and potential isotopic fractionation caused by processing. The lipid standard was a mixture of a wax ester (C_{14} alcohol, C_{20} fatty acid), C_{19} alcohol, C_{19} fatty acid and androstanol. The amino-acid and sugar standards were L-leucine and D-glucose, respectively. Procedural blanks for protein and carbohydrate extraction methods were processed through the hydrolysis, neutralization, and elution steps using only the reagents. Blanks (2–3 each) needed to be combined to yield enough carbon for ^{14}C analyses. No lipid-extract blanks were measured isotopically because of their extremely low ($\leq 1 \mu\text{g}$) carbon yields. Mean standard recoveries were $76 \pm 33\%$ ($n = 4$) for the lipid extract, $66 \pm 12\%$ ($n = 4$) for the protein-like fraction and $93 \pm 17\%$ ($n = 4$) for the carbohydrate-like fraction.

Carbon isotope analyses

Total DOM_{HMW} and organic fractions (lipid extract, protein- and carbohydrate-like) were dried *in vacuo* and acidified with 1% H_3PO_4 overnight to remove any carbonates or dissolved inorganic carbon. Acidified samples were then dried *in vacuo* and the organic carbon oxidized to CO_2 at 850 °C in evacuated sealed quartz tubes containing CuO and Ag metal²⁶. The CO_2 was purified and the yield quantified using a calibrated Baratron absolute pressure gauge (MKS Industries) on a vacuum extraction line. Samples were split into sealed Pyrex tubes for $\Delta^{14}\text{C}$ (~90% of total CO_2) and $\delta^{13}\text{C}$ (~10% of total CO_2) determinations. $\Delta^{14}\text{C}$ is defined here as the ‰ deviation of the $^{14}\text{C}/^{12}\text{C}$ ratio (^{13}C -normalized and corrected to AD 1950) for the sample relative to the $^{14}\text{C}/^{12}\text{C}$ ratio of the absolute international standard (95% of the AD 1950 activity of NBS Oxalic Acid I, normalized to $\delta^{13}\text{C} = -19\text{‰}$)²⁷. Sample CO_2 for $\Delta^{14}\text{C}$ analyses was reduced to elemental graphite using H_2 over Co catalyst²⁸. $\Delta^{14}\text{C}$ analyses were performed by accelerator mass

spectrometry (AMS) and $\delta^{13}\text{C}$ measurements were made using a Finnigan Delta S isotope ratio mass spectrometer. The $\Delta^{14}\text{C}$ and $\delta^{13}\text{C}$ values of standard lipid, amino-acid and sugar compounds following extraction procedures were not significantly different from those of the pure compounds, indicating that both blank and fractionation effects were minimal. Errors ($\pm 1\sigma$) were considered to be the larger of either standard compound analyses or sample replicates ($n = 2-3$), and were $\pm 9\text{‰}$ for DOM_{HMW} , $\pm 49\text{‰}$ for the protein-like fraction and $\pm 75\text{‰}$ for the carbohydrate-like fraction for $\Delta^{14}\text{C}$ measurements, and $\pm 0.9\text{‰}$ for all $\delta^{13}\text{C}$ measurements. Owing to the small sample sizes, replicate lipid extractions were not possible. Errors for lipid-extract $\Delta^{14}\text{C}$ measurements were $\pm 4-7\text{‰}$ based on AMS analytical uncertainties. The ^{14}C ages of Σ DOM from this study were identical within measurement error to those determined previously at these two sites, and had not changed significantly over the 10–15 yr between the first and second occupations of the Pacific and Atlantic stations (J.E.B., unpublished data).

Weighted mean turnover time calculation

Weighted mean turnover times (TOT) for bulk DOM (Σ DOM) having non-homogeneous component ages were estimated by calculating first-order turnover rate constants for each individual DOM_{HMW} organic subcomponent (lipid extract, protein-like, carbohydrate-like and MUC fractions) and DOM_{LMW} as:

$$k_i = \frac{F_i}{A_i}$$

where k_i is the turnover rate constant in yr^{-1} for fraction i , F_i is the relative contribution of fraction i to the Σ DOM pool, and A_i is the ^{14}C age of fraction i (Tables 1 and 2, Supplementary Table S1). The weighted mean TOT²⁵ of Σ DOM is then calculated as:

$$\text{TOT}_{\Sigma\text{DOM}} = \frac{1}{\sum k_i}$$

Received 12 December 2003; accepted 24 June 2004; doi:10.1038/nature02780.

- Williams, P. M. & Druffel, E. R. M. Radiocarbon in dissolved organic carbon in the central North Pacific Ocean. *Nature* **330**, 246–248 (1987).
- Hedges, J. I. Global biogeochemical cycles: progress and problems. *Mar. Chem.* **39**, 67–93 (1992).
- Amon, R. M. W. & Benner, R. Bacterial utilization of different size classes of dissolved organic matter. *Limnol. Oceanogr.* **41**, 41–51 (1996).
- Carlson, C. A. in *Biogeochemistry of Marine Dissolved Organic Matter* (eds Hansell, D. A. & Carlson, C. A.) 91–151 (Academic, Orlando, USA, 2002).
- Benner, R. in *Biogeochemistry of Marine Dissolved Organic Matter* (eds Hansell, D. A. & Carlson, C. A.) 59–90 (Academic, Orlando, USA, 2002).
- Bauer, J. E., Williams, P. M. & Druffel, E. R. M. ^{14}C activity of dissolved organic carbon fractions in the north central Pacific and Sargasso Sea. *Nature* **357**, 667–670 (1992).
- Druffel, E. R. M., Williams, P. M., Bauer, J. E. & Ertel, J. R. Cycling of dissolved and particulate organic matter in the open ocean. *J. Geophys. Res.* **97**, 15639–15659 (1992).
- Stuiver, M., Quay, P. D. & Ostlund, H. G. Abyssal water ^{14}C distribution and the age of the world oceans. *Science* **219**, 849–851 (1983).
- Eglinton, T. I. et al. Composition, age, and provenance of organic matter in NW African dust over the Atlantic Ocean. *Geochim. Geophys. Res.* **3**, doi:10.1029/2001GC000269 (2002).
- Goericke, R., Montoya, J. P. & Fry, B. in *Stable Isotopes in Ecology and Environmental Science* (eds Lajtha, K. & Michener, R. H.) 181–221 (Blackwell Scientific Publications, Oxford, UK, 1994).
- Eglinton, T. I. et al. Variability in radiocarbon ages of individual organic compounds from marine sediments. *Science* **277**, 796–799 (1997).
- Loh, A. N. *Chemical, Isotopic and Microbial Characterization of Dissolved and Particulate Organic Matter in Estuarine, Coastal and Open Ocean Systems*. Doctoral dissertation, College of William and Mary (2002).
- Wang, X.-C., Druffel, E. R. M., Griffin, S., Lee, C. & Kashgarian, M. Radiocarbon studies of organic compound classes in plankton and sediment of the northeastern Pacific Ocean. *Geochim. Cosmochim. Acta* **62**, 1365–1378 (1998).
- Hwang, J. & Druffel, E. R. M. Lipid-like material as the source of the uncharacterized organic carbon in the ocean? *Science* **299**, 881–884 (2003).
- Wakeham, S. G., Hedges, J. I., Lee, C., Peterson, M. L. & Hernes, P. J. Compositions and transport of lipid biomarkers through the water column and surficial sediments of the equatorial Pacific Ocean. *Deep-Sea Res.* **44**, 2131–2162 (1997).
- Aluwihare, L. I., Repeta, D. J. & Chen, R. F. Chemical composition and cycling of dissolved organic matter in the mid-Atlantic bight. *Deep-Sea Res.* **49**, 4421–4437 (2002).
- Hedges, J. I. et al. The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31**, 945–958 (2000).
- Aluwihare, L. I., Repeta, D. J. & Chen, R. F. A major biopolymeric component to dissolved organic carbon in surface seawater. *Nature* **387**, 166–169 (1997).
- Santschi, P. H. et al. Isotopic evidence for the contemporary origin of high-molecular weight organic matter in oceanic environments. *Geochim. Cosmochim. Acta* **59**, 625–631 (1995).
- Druffel, E. R. M., Bauer, J. E., Griffin, S. & Hwang, J. Penetration of anthropogenic carbon into organic particles of the deep ocean. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL017423 (2003).
- Clark, L. L., Ingall, E. D. & Benner, R. Marine phosphorus is selectively remineralised. *Nature* **393**, 426 (1998).
- Masiello, C. A. & Druffel, E. R. M. Black carbon in deep-sea sediments. *Science* **280**, 1911–1913 (1998).
- Wang, X.-C., Chen, R. F., Whelan, J. & Eglinton, L. Contribution of "old" carbon from natural marine hydrocarbon seeps to sedimentary and dissolved organic carbon pools in the Gulf of Mexico. *Geophys. Res. Lett.* **28**, 3313–3316 (2001).
- Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: an assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
- Trumbore, S. E. & Druffel, E. R. M. in *The Role of Nonliving Organic Matter in the Earth's Carbon Cycle* (eds Zepp, R. G. & Sonntag, Ch.) 7–22 (John Wiley & Sons, Chichester, UK, 1995).
- Sofer, Z. Preparation of carbon dioxide for stable carbon isotope analysis of petroleum fractions. *Anal. Chem.* **52**, 1389–1391 (1980).

27. Stuiver, M. & Polach, H. A. Discussion: reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
28. Vogel, J. S., Southon, J. R. & Nelson, D. E. ^{14}C background levels in an AMS system. *Nucl. Instrum. Methods Phys. Res.* **29**, 50–56 (1987).
29. Masiello, C. A., Druffel, E. R. M. & Currie, L. A. Radiocarbon measurements of black carbon in aerosols and ocean sediments. *Geochim. Cosmochim. Acta* **66**, 1025–1036 (2002).
30. Whelan, J. K. & Thompson-Rizer, C. L. in *Organic Geochemistry: Principles and Applications* (eds Engel, M. H. & Macko, S. A.) 289–353 (Plenum, New York, USA, 1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Canuel, J. Hwang and S. Griffin for laboratory guidance during compound class extractions; M. Ederington-Hagy, E. Waterson and J. Southon for discussions on experiments; S. Griffin, R. Wilson, L. Delizo, C. Masiello, A. Grottolli and the captains and crews of RV *Melville* and RV *Knorr* for field assistance and logistical support; A. McNichol and colleagues at NOSAMS for $\delta^{14}\text{C}$ measurements; E. Franks for $\delta^{13}\text{C}$ measurements; and R. Benner for comments that significantly improved this manuscript. This work was supported by the Chemical Oceanography Program of the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.N.L. (anloh@fgcu.edu).

Impact of climate change on marine pelagic phenology and trophic mismatch

Martin Edwards & Anthony J. Richardson

Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

Phenology, the study of annually recurring life cycle events such as the timing of migrations and flowering, can provide particularly sensitive indicators of climate change¹. Changes in phenology may be important to ecosystem function because the level of response to climate change may vary across functional groups and multiple trophic levels. The decoupling of phenological relationships will have important ramifications for trophic interactions, altering food-web structures and leading to eventual ecosystem-level changes. Temperate marine environments may be particularly vulnerable to these changes because the recruitment success of higher trophic levels is highly dependent on synchronization with pulsed planktonic production^{2,3}. Using long-term data of 66 plankton taxa during the period from 1958 to 2002, we investigated whether climate warming signals⁴ are emergent across all trophic levels and functional groups within an ecological community. Here we show that not only is the marine pelagic community responding to climate changes, but also that the level of response differs throughout the community and the seasonal cycle, leading to a mismatch between trophic levels and functional groups.

The vast majority of documented phenology studies relating seasonal shifts in biology to climate have come from terrestrial and limnological sources (see refs 5, 6). Furthermore, most studies have solely reported phenological changes for a single species and have not explored trophic and ecological interactions⁷. In this study we investigated changes in marine pelagic phenology in the North Sea across three trophic levels using five functional groups. The major functional groups included diatoms and dinoflagellates separately (primary producers); copepods (secondary producers); non-copepod holozooplankton (secondary and tertiary producers) and meroplankton including fish larvae (secondary and tertiary producers). Inter-annual changes in a measure of the timing of the seasonal peak throughout the whole pelagic production season

(the central tendency; see Methods and Fig. 1a, b) were calculated using data from the Continuous Plankton Recorder (CPR)⁸, one of the longest and most spatially extensive marine biological data sets in the world.

The x axis of Fig. 1c shows the timing of the seasonal peaks in 1958 of all 66 plankton taxa used in the analysis; this represents the classical view of succession in the temperate marine pelagic ecosystem. Using the linear slope of the time series of the timing of the seasonal peak, we calculated the change in timing of the seasonal cycle (in months) from 1958 to 2002 for each taxon (Fig. 1c; y axis). Substantial temporal modifications in seasonal successional peaks have occurred over the past few decades. In particular, seasonal peaks of meroplankton have moved significantly ($P < 0.0001$) forward (for example, the phylum Echinodermata has moved by 47 days (d)). By contrast, diatom peaks in spring and autumn have collectively remained relatively static, albeit with considerable

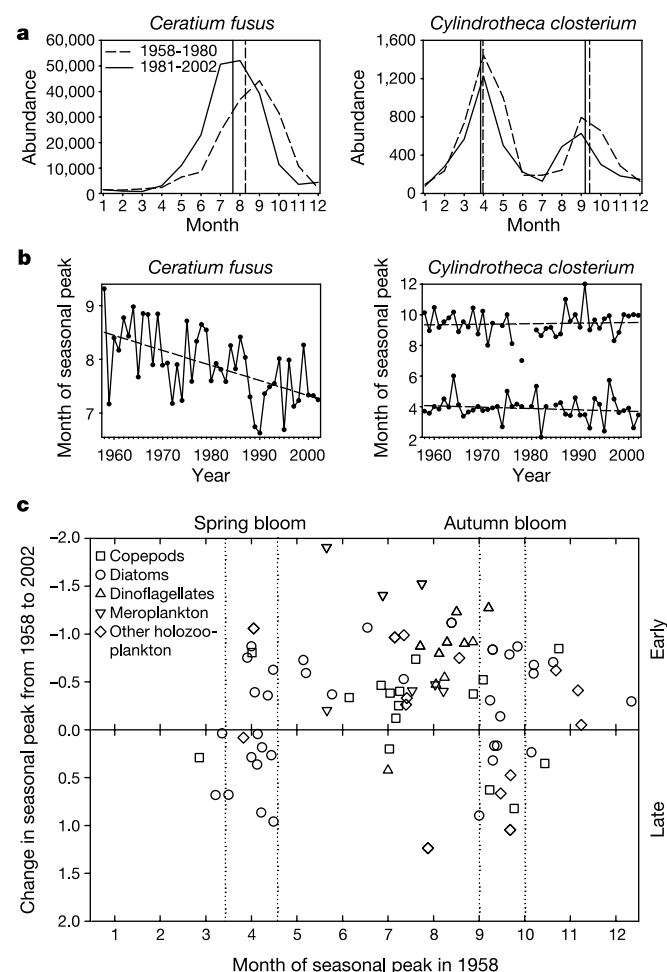


Figure 1 Changes in phenology throughout the pelagic season. **a**, Examples of seasonal cycles for two of the 66 taxa—the dinoflagellate *Ceratium fusus* and the diatom *Cylindrotheca closterium*—used in the analysis for the periods 1958–1980 and 1981–2002. The timing of the seasonal peaks, using the indicator of central tendency, is also shown. **b**, Inter-annual variability of the seasonal peak for the above two species from 1958 to 2002. **c**, The change in the timing of the seasonal peaks (in months) for the 66 taxa over the 45-yr period from 1958 to 2002 plotted against the timing of their seasonal peak in 1958. For each taxon, the linear regression in **b** was used to estimate the difference between the seasonal peak in 1958 and 2002. A negative difference between 1958 and 2002 indicates seasonal cycles are becoming earlier. Standard linear regression was considered appropriate because there was minimal autocorrelation (determined by the Durbin–Watson statistic) in the phenology time series.

inter-taxon variation. Other holozooplankton taxa (mainly of the class Malacostraca) show a wide variety of responses, although the seasonal cycles of the classes Branchiopoda (genus *Evadne*, 30 d) and Appendicularia (30 d), as well as hyperiids (order Amphipoda, class Malacostraca, 32 d) occur substantially earlier. Cycles of the majority of dinoflagellates are also earlier (principally the genera *Ceratium*, 27 d, *Protoperidinium*, 26 d and *Dinophysis*, 24 d), having

important implications for the monitoring and study of harmful algal blooms.

The general pattern observed for taxa that peak when the water column is mixed or in a transitional state is to show considerable variability in phenology, whereas taxa associated with low turbulent conditions have virtually all advanced in their seasonality (34 out of 37 taxa between May–August). The benthic larval component of the zooplankton (meroplankton) has also shown larger shifts forward in seasonality compared with the holozooplankton. During summer, meroplankton have moved forward collectively by 27 d, dinoflagellates by 23 d, copepods by 10 d and non-copepod holozooplankton by 10 d over the 45-yr study period. Diatoms as a group showed the largest variations in phenology, with particular taxa occurring both earlier and later during the spring and autumn blooms. Diatoms show a wide variety of life strategies, and, following the Sverdrup model⁹, delays in the spring bloom for some species could be associated with recent increases in wind forcing in the North Sea and trends in the North Atlantic Oscillation index¹⁰. Collectively, however, the mean movement in the spring bloom was 0 d and the mean movement in the autumn bloom was to 5 d earlier. Many long-term phytoplankton studies have noted that the timing of the spring bloom is in fact fairly constant, occurring approximately the same time each year under highly variable environmental conditions^{11,12}. Other studies have also shown that the development of the pycnocline (which shows considerable geographical/temporal variability in European shelf seas) is not an essential prerequisite for the development of the spring bloom^{13,14}. Recent research has implicated photoperiod in the control of diatom spore growth and germination^{11,12,15}. Thus the temporal stasis in the spring bloom could be a consequence of the diatom community in this study being dominated by taxa that form resting spores.

To explore the relationship between the timing of the seasonal peak and hydro-climatic change we correlated the annual centre of gravity index for the period 1958–2002 for the five functional groups (expressed as principal components) with spring sea surface temperature (SST) (data supplied by Hadley Centre for Climate Research). We adjusted significance levels to account for temporal autocorrelation, and found highly significant correlations between the first principal component (PC1) of the peak in the seasonal cycle of summer plankton and SST (Fig. 2), particularly for dinoflagellates ($r = 0.69$, $P < 0.0001$; PC1 45.5% of the total variability) and meroplankton ($r = 0.70$, $P < 0.0001$; PC1 37.2%). Also during summer, diatoms ($r = 0.31$, $P > 0.05$; PC1 22%), copepods ($r = 0.55$, $P < 0.001$; PC1 33%) and other holozooplankton ($r = 0.34$, $P < 0.05$; PC1 24%) showed positive correlations between phenology and SST. By contrast, diatoms, copepods and other holozooplankton that had peaks in spring and autumn (that is, taxa with bimodal seasonal cycles; data not shown) showed no relationship between phenology and SST. For additional information on a species-by-species basis, individual correlations and their confidence limits are given in the Supplementary Information.

The relationship between SST and the seasonal development of some plankton taxa, particularly the meroplankton and holozooplankton, can be explained by the species-specific effects of temperature on many aspects of plankton physiology, such as adult mortality, reproduction, respiration, embryonic and gonad development^{16–20}. Dinoflagellates may not only be responding physiologically to temperature, but may also respond to temperature indirectly if climate warming enhances stratified conditions and/or if these conditions appeared earlier in the season. Stratified conditions are predicted to intensify with patterns of climate change in the North Sea²¹. The magnitude of changes in phenology reported in this study are greater than those from previous studies on terrestrial communities⁴ and indicate that marine pelagic communities are particularly sensitive to climate change. These large phenological shifts have occurred with an increase in SST of 0.90 °C (estimated from regression) during the study period (1958–2002).

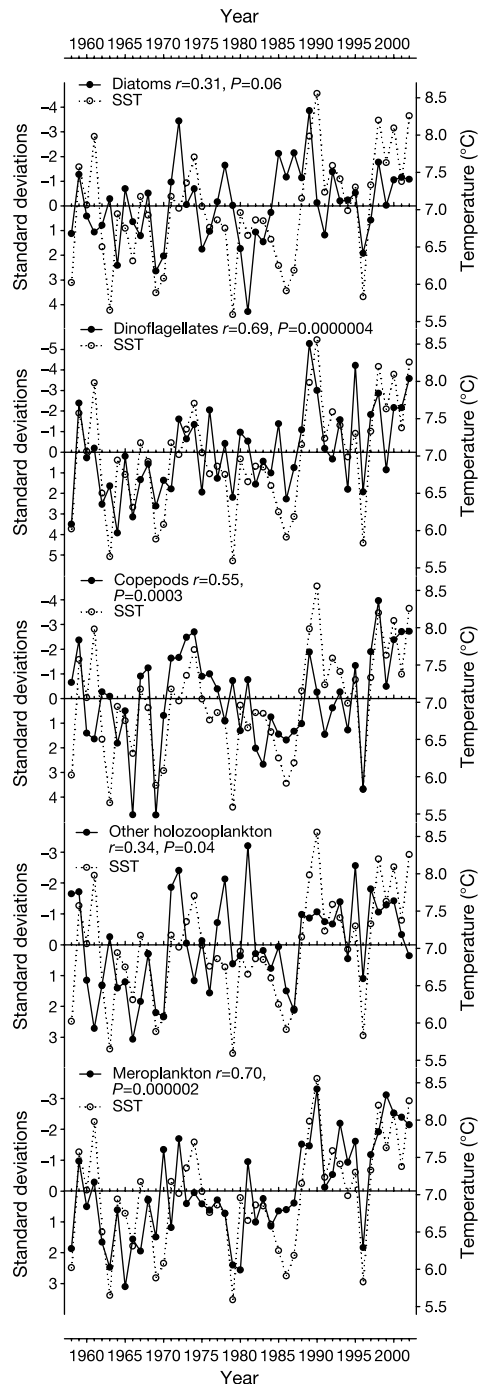


Figure 2 The relationship between the interannual variation in the timing of the seasonal cycle for various functional groups during the summer stratified period and SST. Note the high correlations for dinoflagellates and meroplankton. The time series of the timing of the seasonal cycle for each functional group was represented by principal component analysis of all constituent taxa. Negative standard deviations represent earlier seasonal cycles.

Although we have observed considerable inter-annual variability in plankton phenology, significant underlying patterns over the past few decades have emerged. Diatom blooms in spring, and hence the beginning of the pelagic seasonal cycle, have remained relatively fixed in time, and are presumably dependent on day length or light intensity^{11,15} rather than on temperature-mediated physiological responses in their life strategies. Conversely, organisms that are dependent on temperature to stimulate physiological developments and larval release have significantly moved forward in their seasonal cycle in response to temperature, a trend that has continued over the last decade (with the exception of 1996, a negative North Atlantic Oscillation index year). Although many pelagic organisms are responding to climate warming, it is the intensity of the response that varies considerably amongst the pelagic assemblage. The different extent to which functional groups are moving forward in time in response to warming has led to mismatch between successive trophic levels and a change in the synchrony of timing between primary, secondary and tertiary production. Because efficient transfer of marine pelagic production to higher trophic levels, such as commercially important fish species, is largely dependent on the temporal synchrony between successive trophic production peaks^{2,3}, our study suggests that marine trophodynamics may have already been radically altered (notwithstanding some species adaptations), and will continue to do so in the coming decades if the climate continues to warm at its present rate. In addition to the effects of overfishing, the decline in abundance of key planktonic prey, and shifts in their seasonality, have recently been implicated in exacerbating the decline in North Sea cod stocks²². Planktonic phenological shifts of the magnitude reported in this study, coupled with large-scale shifts in plankton biogeography²³, will undoubtedly have a considerable effect on ecosystem function in the North Sea and may help to explain the recently reported dramatic ecological changes observed in the North Sea²⁴. □

Methods

Plankton data

The CPR survey is the longest running, large-scale marine biological survey in the world. The CPR is a near-surface (10 m) plankton sampler voluntarily towed each month behind merchant ships on their normal routes of passage. Methods of analysis for ~400 phyto- and zooplankton taxa have remained almost unchanged since 1958⁸. In this study we use data from the most consistently sampled region in the entire survey, the central North Sea (55–58° N). From 1958 to 2002 all months have been sampled in this region. Species found in more than 1% of the samples were included in the analysis, as variations each year in the seasonal cycle of rarer species are not adequately estimated. To aid interpretation, species were assigned to functional groups, namely diatoms, dinoflagellates, copepods, non-copepod holozooplankton and meroplankton.

Sea surface temperature

SST was used as an indicator of climate change in the ocean because organisms respond physiologically to temperature and it has been found to be important in many terrestrial phenological studies¹. Monthly mean gridded (1° × 1°) SSTs from a blend of satellite AVHRR (advanced very high resolution radiometer) and *in situ* observations were obtained from the Hadley Centre of the UK Met Office (HadISST). We calculated mean spring SST in the central North Sea, because this time of year is important for the seasonal cycles of many pelagic organisms. Similar results, however, were obtained for mean annual SST (data not shown).

Data analysis

Most terrestrial phenological studies have used the first arrival (for example, of migratory species such as swallows) or leaf emergence date (for example, oaks) of a particular species to describe phenological changes²⁵. In the present study, we estimated the timing of the seasonal peak throughout the entire growing season (the central tendency, *T*) using the month co-ordinate of the centre of gravity of the area below graphs of monthly means^{26,27}:

$$T = \frac{\sum_{m=1}^{12} M x_m}{\sum_{m=1}^{12} x_m}$$

where x_m is the mean abundance in month *M* (January = 1, ..., December = 12).

This index is sensitive to changes in the timing of the seasonal cycle (Fig. 1; see also refs 26, 27). The average seasonal cycle over the 45-yr period for each taxon was used to determine whether taxa were unimodal or bimodal (spring and autumn). A fundamental difference between terrestrial and many pelagic ecosystems is that there is just one seasonal primary production peak in terrestrial ecosystems but usually two in temperate marine

environments: one in spring and the other in autumn. During autumn, the pelagic environment reverts back to spring-like conditions, accompanied by another burst in phytoplankton production, albeit not as intensive as the spring bloom. As a result, many pelagic organisms, in particular the floral community (predominantly the class Bacillariophyceae, the diatoms, that are nutrient limited during the summer period), display bimodal seasonality. For unimodal taxa the timing of the seasonal peak was calculated throughout the entire year, whereas for bimodal taxa the timing of the seasonal peak was calculated separately for the first six months and the last six months of the year.

To summarize phenology time series of the different species within each functional group, standardized principal components analysis (PCA) based on a correlation matrix was used²⁸. These PCAs based on phenology were correlated with mean spring SST and were adjusted for temporal autocorrelation using the modified Chelton method²⁹.

Potential biases

A possible source of bias with the central tendency index is associated with changes in the time each month that sampling takes place. Six routes are generally towed within the central North Sea, which helps to minimize bias, although mean sampling time each month has still moved forward by an average of 2.4 d over the entire period from 1958 to 2002. This is relatively small compared with the large shifts earlier of dinoflagellates (23 d) and meroplankton (27 d) over the 45-yr study (Fig. 1). In addition, the slightly earlier sampling each month, over more recent times, would actually tend to delay and not bring forward the timing of peak abundance as calculated by *T*, because the seasonal cycle when calculated on a monthly basis would shift to the right. To minimize the likelihood of committing type I errors when identifying changes in the timing of the seasonal cycle for each taxon over the 45-yr study period, we used a conservative significance level of 1%. For some taxa that are not speciated (for example, fish larvae) it is difficult to separate observed trends in phenology from changes in community composition caused by climate-induced biogeographical shifts. However, on the basis of the strong and consistent relationships observed between these taxa and SST, coupled with similar patterns in the majority of speciated taxa, we conclude that many components of the pelagic assemblage are responding rapidly to temperature change through changes in the timing of their seasonal cycles.

Received 9 March; accepted 29 June 2004; doi:10.1038/nature02808.

- Hughes, L. Biological consequences of global warming: is the signal already apparent? *Trends Ecol. Evol.* **15**, 56–61 (2000).
- Hjort, J. Fluctuations in the great fisheries of northern Europe. *Rapp. Conserv. Explor. Mer* **20**, 1–228 (1914).
- Cushing, D. H. Plankton production and year-class strength in fish population: an update of the match/mismatch hypothesis. *Adv. Mar. Biol.* **26**, 250–293 (1990).
- Root, T. L. *et al.* Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
- Walther, G. *et al.* Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002).
- Parnes, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
- Harrington, R., Woiwod, I. & Sparks, T. Climate change and trophic interactions. *Trends Ecol. Evol.* **14**, 146–150 (1999).
- Warner, A. J. & Hays, G. C. Sampling by the Continuous Plankton Recorder survey. *Prog. In Oceanogr.* **34**, 237–256 (1994).
- Sverdrup, H. U. On conditions for the vernal blooming of phytoplankton. *J. Cons. Perm. Int. Explor. Mer.* **18**, 287–295 (1953).
- Siegmund, F. & Schrum, C. Decadal changes in wind forcing over the North Sea. *Clim. Res.* **18**, 39–45 (2001).
- Eilertsen, H. C., Sandberg, S. & Tollefsen, H. Photoperiodic control of diatom spore growth; a theory to explain the onset of phytoplankton blooms. *Mar. Ecol. Prog. Ser.* **116**, 303–307 (1995).
- McQuoid, M. R. & Hobson, L. A. Diatom resting stages. *J. Phycol.* **32**, 889–902 (1996).
- Townsend, D. W., Keller, M. D., Sieracki, M. E. & Ackleson, S. G. Spring phytoplankton blooms in the absence of vertical water column stratification. *Nature* **360**, 59–62 (1992).
- van Haren, H., Mills, D. K. & Wetsteyn, L. P. M. J. Detailed observations of the phytoplankton spring bloom in the stratifying central North Sea. *J. Mar. Res.* **56**, 655–680 (1998).
- Eilertsen, H. C. & Wyatt, T. Phytoplankton models and life history strategies. *S. Afr. J. Mar. Sci.* **22**, 323–338 (2000).
- Greve, W. in *Phenology: An Integrative Environmental Science* (ed. Schwartz, M. D.) 385–403 (Kluwer, Netherlands, 2004).
- Wear, R. G. Incubation in British decapod Crustacea and the effects of temperature on the rate and success of embryonic development. *J. Mar. Biol. Assoc. UK* **54**, 745–762 (1974).
- Lindley, J. A. Diversity, biomass and production of decapod crustacean larvae in a changing environment. *Invertebr. Reprod. Dev.* **33**, 209–219 (1998).
- Hirst, A. G. & Bunker, A. J. Growth of marine planktonic copepods: global rates and patterns in relation to chlorophyll *a*, temperature, and body weight. *Limnol. Oceanogr.* **48**, 1988–2010 (2003).
- Greve, W., Reiners, F. & Nast, J. Biocenotic changes of the zooplankton in German Bight: the possible effects of eutrophication and climate. *ICES J. Mar. Sci.* **53**, 951–956 (1996).
- Van Beusekom, J. E. E. & Diel-Christensen, S. *North Sea Quality Status Report* 64–71 (Danish Environmental Protection Agency, Copenhagen, 1994).
- Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S. & Reid, P. C. Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661–664 (2003).
- Beaugrand, G., Reid, P. C., Ibanez, F., Lindley, J. A. & Edwards, M. Reorganisation of North Atlantic marine copepod biodiversity and climate. *Science* **296**, 1692–1694 (2002).
- Schiermeier, Q. Climate findings let fishermen off the hook. *Nature* **428**, 4 (2004).
- McCarty, J. P. Ecological consequences of recent climate change. *Conserv. Biol.* **15**, 320–331 (2001).
- Colebrook, J. M. & Robinson, G. A. Continuous Plankton Records: seasonal cycles of phytoplankton and copepods in the north-eastern Atlantic and the North Sea. *Bull. Mar. Ecol.* **6**, 123–139 (1965).

27. Colebrook, J. M. Continuous Plankton Records: seasonal cycles of phytoplankton and copepods in the north Atlantic Ocean and the North Sea. *Mar. Biol.* **51**, 23–32 (1979).
28. Beaugrand, G., Ibanez, F. & Lindley, J. A. An overview of statistical method applied to the CPR data. *Prog. In Oceanogr.* **58**, 235–262 (2003).
29. Pyper, B. J. & Peterman, R. M. Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Can. J. Fish. Aquat. Sci.* **55**, 2127–2140 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements A funding consortium made up of governmental agencies from Canada, France, Iceland, Ireland, the Netherlands, Portugal, the UK and the USA financially supports the CPR survey. Main support for this work was provided by UK DEFRA and UK NERC. We would also like to thank J. Bishop, K. Brander, B. Clarke, R. Harris, R. Myers, D. Schoeman and A. Walne for comments on the manuscript and the owners and crews of the ships that tow the CPRs on a voluntary basis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.E. (maed@sahfos.ac.uk).

Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating

Susan Kalisz¹, Donna W. Vogler^{1,2} & Kristen M. Hanley¹

¹University of Pittsburgh, Department of Biological Sciences, Pittsburgh, Pennsylvania 15260, USA

²State University of New York, College at Oneonta, Biology Department, Oneonta, New York 13820, USA

The evolution of self-fertilization in hermaphrodites is opposed by costs that decrease the value of self progeny relative to that of outcross progeny^{1–3}. However, self-fertilization is common in plants⁴; 20% are highly selfing and 33% are intermediate between selfing and outcrossing⁵. Darwin⁶ proposed an adaptive benefit of self-pollination in providing reproductive assurance when outcrossing is impossible^{6–9}. Moreover, if outcross pollen receipt is inconsistent within or between years, these conditions likewise favour self-pollination¹⁰, and this can result in a mixture of self and outcross seed production (mixed mating). Despite wide acceptance, the reproductive assurance hypothesis has lacked the support of complete empirical evidence to show that variable pollination can create both the ecological and genetic conditions favouring self-pollination. We recently showed in *Collinsia verna* that during periods of infrequent pollinator visits, autonomous self-pollination boosted seed output per flower¹¹, the key ecological condition. Here we show low inbreeding depression and marker-based estimates of selfing, demonstrating that when the pollination environment in wild populations necessitates reproductive assurance, selfing rates increase. We provide a complete demonstration of reproductive assurance under variable pollination environments and mechanistically link reproductive assurance to intermediate selfing rates through mixed mating.

Populations of flowering plants that lack mates or pollinators, such as those at the edge of a species' range or colonizing species, rapidly evolve autonomous self-fertilization^{12,13} (within-flowers selfing without a pollen vector)⁸, and this is thought to occur because selfing provides reproductive assurance^{6–9}. Other ecological factors, such as unpredictable outcross pollen receipt within or

among years^{8,10,12,14} may also, in theory, produce conditions that favour autonomous selfing through reproductive assurance^{10,15,16}. For autonomous selfing to evolve, its benefits must be balanced against the potential costs. Alleles promoting self-fertilization ability increase in frequency because plants that carry them can serve as pollen parents in two ways: both by fertilizing ovules on other plants (that is, outcrossing) and by fertilizing their own ovules (that is, selfing)¹⁷. Thus, individuals with alleles that cause more selfing have an advantage in transmission over individuals with alleles for outcrossing¹⁸. In contrast, selfing is disfavoured when there is inbreeding depression (δ , low vigour of self progeny)^{1,2} and/or when the production of selfed progeny pre-empts the production of outcrossed progeny (pollen or seed discounting)^{8,16}. Previous investigations have failed to show that when outcross pollen receipt is inconsistent, selfing is favoured and outweighs these costs¹⁹. We are currently unable to predict when autonomous self-fertilization will provide reproductive assurance. An unequivocal demonstration of reproductive assurance under unreliable pollinators requires several types of data^{4,14,19}. Plants must fail to receive outcross pollen, but this failure need not occur every season. During periods of low or no outcross pollen receipt, autonomous selfing must boost seed production. Finally, the combined costs (seed discounting, pollen discounting and inbreeding depression) must not completely negate the fitness gain of selfing.

Costs incurred by autonomous selfing vary depending both upon the timing of self-pollination relative to outcross pollen receipt⁸ and the availability of pollinators. When pollinators are present, autonomous self-pollination that occurs after all opportunities for outcross pollen receipt have passed (delayed selfing) incurs no pollen or seed discounting costs^{8,10}. Additionally, even if inbreeding depression (δ) is high, the survival of any progeny produced by delayed selfing always provides reproductive assurance¹⁰. In contrast, if autonomous self-pollination coincides with outcross pollen receipt (competing selfing), then pollen discounting, seed discounting and inbreeding depression can disfavour selfing⁸. In theory, if the fitness of self progeny produced by competing selfing is less than or equal to roughly half the fitness of outcrossed progeny (that is $\delta > 0.5$), then the fitness gain due to the transmission advantage is lost^{1,2}. Finally, when pollinators are absent, there are no seed and pollen discounting costs of autonomous selfing^{3,8,10,20}. In a field experiment, we previously investigated autonomous self-pollination in the winter annual wildflower, *C. verna* (Plantaginaceae). We showed that this species autonomously self-pollinates²¹ in a field experiment that compared fruit set of emasculated versus control flowers (Tables 1 and 2 in ref. 11). Further, we quantified the timing of autonomous selfing by comparing both the timing of pollen deposition and the number of pollen grains on the stigmas of flowers in open-pollinated conditions relative to flowers in pollinator-excluded treatments. Selfing in *C. verna* is autonomous and predominantly delayed, with the potential for some competing selfing (Fig. 2 in ref. 11).

Because competing selfing can occur in *C. verna*, it is important to estimate the magnitude of inbreeding depression. Here we report results from three wild populations (BT, EF and TMC; see Methods) located in southwestern Pennsylvania, USA. We produced both selfed and outcrossed progeny on plants from each population and compared their lifetime performance. Mean trait values of self versus outcross progeny for each population were compared, and indicate that only one of the 15 comparisons was statistically significant (seed weight; BT population, $P < 0.001$). All three populations show markedly low average levels of inbreeding depression (Fig. 1; $\delta < 0.15$ for all traits measured, in all populations), lower than the ~ 0.5 value that opposes competing selfing. Additionally, previous field estimates of early inbreeding depression in the three study populations revealed no significant difference (for all six comparisons $P > 0.2$) in the fruiting success of selfed versus

outcrossed flowers¹¹. Our current results are consistent with both field estimates of δ (Black, B., unpublished data) and previous greenhouse studies of populations from Illinois and Michigan, USA, that used identical methods to quantify inbreeding depression²² (and unpublished data, S.K., Holtsford, Thiede, Kärkkäinen and Black). Moreover, in all our populations, δ estimates for lifetime fitness (R_0 = survival $\times$ flower number) are significantly below 0.5² (Population $R_0\delta$ (s.e.m.): BT, $R_0\delta$ = 0.006 (0.075); EF, $R_0\delta$ = 0.176 (0.058); TMC, $R_0\delta$ = 0.086 (0.099)). Because selfing in *C. verna* is mostly delayed¹¹, pollen and seed discounting costs are negligible across all pollinator environments, even for the occasional flowers with competing selfing¹⁰.

Variation in the pollination environments of *C. verna* was previously quantified in a three-year field study in the same populations¹¹. We compared fruiting success of open-pollinated, emasculated flowers versus outcrossed intact control flowers that were also open to pollinators (Table 2 in ref. 11) and showed that pollinator visits varied significantly (12 periods had significant pollinator failure ($P < 0.002$) leading to decreased seed production; 15 periods had no significant pollinator failure ($P < 0.06$) and no decrease in seed production) within and among flowering seasons and that this low visitation resulted in significantly reduced fruit set of emasculated flowers (range 0–35% less fruit). Most importantly, we compared the fruiting success of emasculated flowers that were open to pollinators with that of intact flowers that were open to pollinators (Table 2 in ref. 11), to show that during those periods of scarce pollinators autonomous selfing augmented fruit set (up to +30%)¹¹, the key ecological condition. Thus, ecological circumstances of reproductive assurance—variation in pollinator service and increase in seed set in the absence of pollinators—exist in these *C. verna* populations.

The selfing rate of a population quantifies the proportion of progeny produced through self-pollination²³ and provides an integrated measure of the realized mating system during one flowering season. All else being equal, selfing rates in *C. verna* should mirror the stochasticity of pollinator visitation and increase when pollinator visitation is low because, as we previously showed, proportionately more seeds are sired through autonomous selfing when pollinators fail to visit flowers¹¹. However, self-pollination and mixed mating can occur even when pollinators are abundant if pollinators transport self-pollen within a flower (facilitated selfing),

or among flowers on the same plant (geitonogamy)⁸. In this way, seeds are sired by a combination of pollinator-mediated movement of outcross pollen, and both pollinator-mediated selfing (geitonogamy + facilitated selfing) and autonomous selfing. When pollinators are abundant, all the above types of selfing suffer the costs of seed and pollen discounting⁸, and selfing is disfavoured^{14,24}. Conversely, when pollinators are sometimes absent and early inbreeding depression values are low, as in our study, the annual seed production will be a mixture of autonomous selfing and outcrossing. In this case, flowers unvisited by pollinators suffer no discounting costs, mixed mating is favoured and intermediate selfing rates are expected.

Here we test the relationship between pollinator failure rate and selfing rate over two years (1999 and 2000) in the same three field populations in which we estimated inbreeding depression. To quantify pollinator failure rates, we compare the proportion of open-pollinated, emasculated flowers that failed to set fruit, with the fruit set failure of paired, hand-pollinated control flowers in both years. None of the study populations experienced significant pollinator failure in 1999. In contrast, both the BT and EF populations experienced significant pollinator failure in 2000, whereas the TMC population again did not. Concordant with expectations, selfing rates vary significantly among populations and years

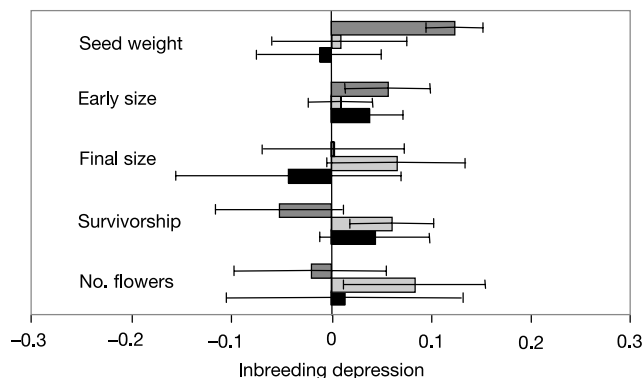


Figure 1 Populations of the annual species, *C. verna*, have low values of inbreeding depression, δ . Mean levels of δ for all populations and traits are <0.15 , below the theoretical value of 0.50 that disfavours competing selfing. Traits were measured across the entire lifespan of the plants. Population-level mean and s.e.m. of δ for each trait were calculated by averaging (family-level difference in trait value between selfed and outcrossed progeny/trait value of the outcrossed progeny). Populations: BT, dark grey; EF, light grey; TMC, black (± 1 s.e.m.). Population differences in mean δ were non-significant (one-way analysis of variance for each trait, $P > 0.2$).

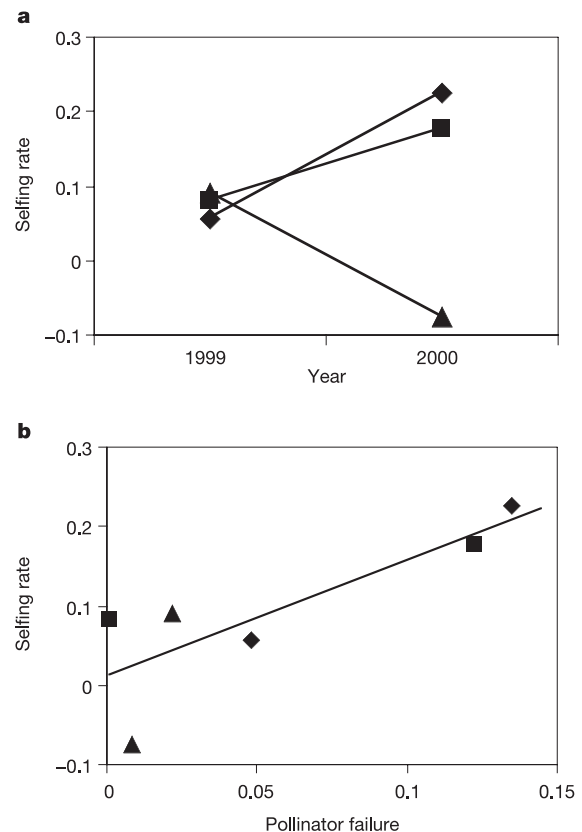


Figure 2 Annual variation in selfing rates in *C. verna* populations results from an increase in autonomous self pollination when pollinators fail to visit flowers. **a**, Selfing rates vary significantly among populations and years. Annual mean selfing rates for each population (BT, diamond; EF, square; TMC, triangle) were determined using multi-locus data. Bootstrapped s.e.m. values are all less than ± 0.01 and are hidden by the symbols. **b**, Selfing rates increase as pollinator visitation rates decrease because more seeds are sired by autonomous selfing. Population selfing rates are significantly and positively correlated with field estimates of pollinator failure ($P < 0.05$, $R^2 = 0.67$, $y = 1.46 + 0.01$). Bootstrapped s.e.m. values are all less than ± 0.01 and are hidden by the symbols.

(Fig. 2a). More importantly, annual selfing rates are positively correlated with pollinator failure rate (Fig. 2b). Our data suggests that a 10% reduction in pollinator visits leads to a 14% increase in the selfing rate. Because only autonomous selfing can occur when pollinators fail to visit a flower, our data mechanistically link reproductive assurance to intermediate selfing through context-dependent changes in the proportion of seed produced by mixed mating.

Our data provide a complete demonstration of reproductive assurance under unpredictable pollinator environments⁴ and an empirical demonstration that selfing rates of populations directly respond to the pollinator environment (Fig. 2). These results indicate that when pollinator visits decrease, populations shift towards intermediate selfing rates through an increase in the proportion of autonomously selfed seeds. This field study provides only a snapshot of the evolutionary process of *C. verna*'s mating system, but brings into focus surprising results. Outcross pollination events generated the majority of the seeds in all populations over both years (73–100% outcrossing, Fig. 2a). However, these selfed progeny express lower early inbreeding depression (Fig. 1) than is typically seen in outcrossing species²⁵, and the magnitude of inbreeding depression increases from early to late stages, as is typically seen in selfing species^{22,25}. Unpredictable pollination environments are the norm in wild plant populations, and can include periods of total pollinator failure^{12,26}. Within the evolutionary history of these *C. verna* populations, years of complete pollinator failure must undoubtedly have occurred. Such extreme pollination environments can both favour autonomous self-pollination^{8,12,13} and reduce genetic load^{25,27}. Species with floral developmental mechanisms that promote outcrossing when pollinators are present, but ensure self-pollination if they are not^{8,11}, can have different annual selfing rates as a functional response to pollinator environments, assuring reproduction and providing a 'best of both worlds' mating system¹². Mating system models have shown that an intermediate level of selfing can be evolutionarily stable^{10,28}. The observation of intermediate selfing rates in many other animal-pollinated species^{5,29} may in part reflect similar fluctuating, context-dependent benefits to selfing and outcrossing in a variable world. □

Methods

Study species and populations

C. verna is a self-compatible winter annual herb, native to the eastern half of North America, that flowers with the spring ephemeral flora. In this species, autonomous self-pollination occurs relatively late in a flower's lifetime and ranges from competing (coincident with outcross pollen receipt) to delayed (after the opportunity for outcrossing has passed)¹¹. Approximately 4–5 days after anthesis begins, the style elongates and brings the receptive stigma into contact with the pollen-bearing anthers³⁰, which can result in self-pollination. The three populations of *C. verna* used in this study are located in different counties and/or watersheds in Pennsylvania, USA: Braddock Trail (BT), Ohio River watershed (Westmoreland Co.); Enlow Fork (EF), Monongahela River watershed (Washington Co.); and Ten Mile Creek (TMC), Ohio River watershed (Washington Co.). These populations differ significantly in their pollinator communities and their dates of first flowering differ by as much as 20 days¹¹. Pollinators include native bumble bees and solitary bees as well as the introduced European honey bee¹¹.

Inbreeding depression

At the end of the 1999 flowering season, 30 plants bearing seeds were collected in the field from each population. All seeds were individually planted and placed under growth chamber conditions that cue germination. One seedling was randomly chosen from each of the original 30 plants, per population, until a sample size of 20 parents was achieved for each population. These parents were grown to flowering in a greenhouse. Eight flowers per parent were emasculated at the bud stage. Four of the emasculated flowers were hand self-pollinated with pollen from other flowers on the same parent plant, whereas the remaining four flowers were outcross-pollinated with a pollen mixture from three to six donor plants from the same population in the experiment. As in a previous study on inbreeding depression in *C. verna*²², the resulting seeds were individually weighed, and six self-pollinated seeds and six outcrossed seeds per parent were germinated in a Conviron controlled environment growth chamber, transplanted and grown to maturity in a greenhouse. Imbalances in the number of progeny per parent reduced the number of parents that could be used in the analyses to 16–18 per population. Traits measured on all progeny were: seed weight (mg), early size (cotyledon diameter, mm), final size (number

of branches + number of whorls at flowering), survival to flowering, number of flowers and lifetime reproductive output (R_0). The difference in mean trait value (fitness) between self-pollinated progeny (w_s) and outcrossed progeny (w_o) estimates the magnitude of inbreeding depression (δ values) calculated as in ref. 2: $\delta = (w_o - w_s)/w_o$. Family-level δ were determined and used to calculate the population-level mean and standard error in δ (Fig. 1). Population mean δ values were compared using a one-way analysis of variance for each trait. Finally, we compared the mean traits values for each population using selfed and outcrossed family means in a two-tailed, paired *t*-test.

Pollinator failure rate and selfing rate

In all three populations in 1999 and 2000, groups of 200 flowers were emasculated at the bud stage and paired with groups of 200 intact flowers. Pollinators do not discriminate against emasculated flowers of *C. verna*³⁰. Four to five days later, intact flowers were hand outcrossed with pollen from three to six pollen donors located at least 1 m away, whereas emasculated flowers received pollen only from natural pollinator visits (open-pollinated). This experiment was repeated three times across the peak flowering period each year ($N = 600$ flowers per treatment per population per year). We define pollinator failure rate as: $1 - (\% \text{ fruit set of emasculated flowers}) / \% \text{ fruit set of hand outcrossed intact flowers}$.

To determine the selfing rate, all seeds from 50 randomly chosen plants per population were placed under germination conditions in growth chambers and 20–35 maternal sibships per population with at least five seedlings were used to determine annual selfing rate per population. Tissue was ground and electrophoresed on an 11% starch gel and stained for seven polymorphic enzyme systems: MDH1, MDH2, UGP1, UGP2, UGP3, DIA, and PGI. Maximum likelihood estimation techniques with 500 bootstraps using family resampling methods were used to calculate the mean and standard error of the selfing rate for each population and year using the MLTR program²³.

Received 5 May; accepted 22 June 2004; doi:10.1038/nature02776.

1. Jarne, P. & Charlesworth, D. The evolution of the selfing rate in functionally hermaphroditic plants and animals. *Annu. Rev. Ecol. Syst.* **24**, 441–466 (1993).
2. Lande, R. & Schmske, D. W. I. Genetic models. *Evolution* **39**, 24–40 (1985).
3. Holsinger, K. E. Mass action models of plant mating systems: the evolutionary stability of mixed mating systems. *Am. Nat.* **138**, 606–622 (1991).
4. Barrett, S. C. H. The evolution of plant sexual diversity. *Nature Rev. Genet.* **3**, 274–284 (2002).
5. Vogler, D. W. & Kalisz, S. Sex among the flowers: the distribution of plant mating systems. *Evolution* **55**, 202–204 (2001).
6. Darwin, C. R. *The Effects of Cross and Self-Fertilization in the Vegetable Kingdom* Ch. 9 (John Murray, London, 1876).
7. Stebbins, G. L. *Flowering Plants: Evolution Above the Species Level* 52 (Belknap, Cambridge, Massachusetts, 1974).
8. Lloyd, D. G. II. The selection of self-fertilization. *Int. J. Plant Sci.* **153**, 370–380 (1992).
9. Baker, H. G. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**, 347–348 (1955).
10. Schoen, D. J. & Brown, A. H. D. Whole- and part-flower self-pollination in *Glycine clandestina* and *G. argyrea* and the evolution of autogamy. *Evolution* **45**, 1665–1674 (1991).
11. Kalisz, S. & Vogler, D. W. Benefits of autonomous selfing under unpredictable pollinator environments. *Ecology* **84**, 2928–2942 (2003).
12. Cruden, R. W. & Lyon, D. L. in *The Evolutionary Ecology of Plants* (eds Bock, J. H. & Linhart, Y. B.) 171–207 (Westview, Boulder, Colorado, 1989).
13. Fausto, J. A. J., Eckhart, E. V. & Geber, M. A. Reproductive assurance and the evolutionary ecology of self-pollination in *Clarkia xantiana* (Onagraceae). *Am. J. Bot.* **88**, 1794–1800 (2001).
14. Barrett, S. C. H. Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Phil. Trans. R. Soc. Lond. B* **358**, 991–1004 (2003).
15. Lloyd, D. G. & Schoen, D. J. I. Functional dimensions. *Int. J. Plant Sci.* **153**, 358–369 (1992).
16. Schoen, D. J., Morgan, M. T. & Bataillon, T. How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Phil. Trans. R. Soc. Lond. B* **351**, 1281–1290 (1996).
17. Holsinger, K. Reproductive systems and evolution in vascular plants. *Proc. Natl Acad. Sci. USA* **97**, 7037–7042 (2000).
18. Fisher, R. A. Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63 (1941).
19. Herlihy, C. R. & Eckert, C. G. Genetic cost of reproductive assurance in a self-fertilizing plant. *Nature* **416**, 320–323 (2002).
20. Holsinger, K. E. Inbreeding depression and the evolution of plant mating systems. *Trends Ecol. Evol.* **6**, 307–308 (1991).
21. Schoen, D. J. & Lloyd, D. G. III. Methods for studying modes and functional aspects of self-fertilization. *Int. J. Plant Sci.* **153**, 381–393 (1992).
22. Kalisz, S. Fitness consequences of mating system, seed weight and emergence date in a winter annual. *Evolution* **43**, 1263–1272 (2002/1989).
23. Ritland, K. Multilocus mating system program MLTR 2.2. (<http://genetics.forestry.ubc.ca/ritland/programs.html>)
24. Harder, L. D. & Barrett, S. C. H. Mating cost of large floral displays in hermaphrodite plants. *Nature* **373**, 512–515 (1995).
25. Husband, B. C. & Schmske, D. W. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* **50**, 54–70 (1996).
26. Burd, M. Bateman's principle and plant reproduction: the role of pollen limitation in fruit and seed set. *Bot. Rev.* **60**, 83–139 (1994).
27. Byers, D. L. & Waller, D. M. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annu. Rev. Ecol. Syst.* **30**, 479–513 (1999).
28. Johnston, M. O. Evolution of intermediate selfing rates in plants: pollination ecology versus deleterious mutations. *Genetica* **102/103**, 267–278 (1998).
29. Barrett, S. C. H., Harder, L. D. & Worley, A. C. The comparative biology of pollination and mating in flowering plants. *Phil. Trans. R. Soc. Lond. B* **351**, 1271–1280 (1996).
30. Kalisz, S. *et al.* The mechanism of delayed selfing in *Collinsia verna* (Scrophulariaceae). *Am. J. Bot.* **86**, 1239–1247 (1999).

Acknowledgements We thank J. Dunn for technical assistance and J. Dunn, P. Zemrowski, A. Richter, C. Jarzab, H. Lang, R. Brown and A. Mergenthaler for field assistance. D.W. Schemske and S. J. Tonsor provided valuable comments on the manuscript. This work was supported by research grants from the National Science Foundation (USA) and the Research Development Fund of The University of Pittsburgh (S.K.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K. (kalisz@pitt.edu).

A barley cultivation-associated polymorphism conveys resistance to powdery mildew

Pietro Piffanelli^{1*†}, Luke Ramsay^{2*}, Robbie Waugh²,
Abdellah Benabdelmouna³, Angélique D'Hont³, Karin Hollricher⁴,
Jørgen Helms Jørgensen⁵, Paul Schulze-Lefert⁶ & Ralph Panstruga⁶

¹The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

²Genomics Unit, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

³CIRAD, Avenue Agropolis, 34398 Montpellier, Cedex 5, France

⁴Max-Planck-Institut für Züchtungsforschung, Department of Plant Breeding and Yield Physiology, Carl-von-Linne-Weg 10, D-50829 Köln, Germany

⁵Risø National Laboratory, Plant (formerly: Agricultural) Research Department, DK-4000 Roskilde, Denmark

⁶Max-Planck-Institut für Züchtungsforschung, Department of Plant-Microbe Interactions, Carl-von-Linne-Weg 10, D-50829 Köln, Germany

* These authors contributed equally to this work

† Present address: CIRAD, Avenue Agropolis, 34398 Montpellier, Cedex 5, France

Barley (*Hordeum vulgare*) has played a pivotal role in Old World agriculture since its domestication about 10,000 yr ago¹. Barley plants carrying loss-of-function alleles (*mlo*) of the *Mlo* locus are resistant against all known isolates of the widespread powdery mildew fungus². The sole *mlo* resistance allele recovered so far from a natural habitat, *mlo*-11, was originally retrieved from Ethiopian landraces and nowadays controls mildew resistance in the majority of cultivated European spring barley elite varieties². Here we use haplotype analysis to show that the *mlo*-11 allele probably arose once after barley domestication. Resistance in *mlo*-11 plants is linked to a complex tandem repeat array inserted upstream of the wild-type gene. The repeat units consist of a truncated *Mlo* gene comprising 3.5 kilobases (kb) of 5'-regulatory sequence plus 1.1 kb of coding sequence. These generate aberrant transcripts that impair the accumulation of both *Mlo* wild-type transcript and protein. We exploited the meiotic instability of *mlo*-11 resistance and recovered susceptible revertants in which restoration of *Mlo* function was accompanied by excision of the repeat array. We infer *cis*-dependent perturbation of transcription machinery assembly by transcriptional interference in *mlo*-11 plants as a likely mechanism leading to disease resistance.

Barley *Mlo* encodes the prototype of a plant-specific family of seven-transmembrane domain proteins^{3–5}. The protein interacts with the Ca²⁺ sensor calmodulin and seems to inhibit a vesicle-associated and SNARE-protein-dependent resistance to the barley powdery mildew fungus (*Blumeria graminis* f. sp. *hordei*; *Bgh*) at the cell periphery^{6–8}. Each of 17 molecularly characterized *mlo* mutants was derived from chemical-induced or radiation-induced mutagenesis, invariably affecting coding or intron splice junction

sequences^{3,9}. Some primitive cultivars (landraces) collected from the granaries of local farmers in Ethiopia during expeditions in 1937 and 1938 possess strong resistance against all tested *Bgh* isolates^{2,10} that genetic analysis has attributed to the presence of *mlo* alleles (designated *mlo*-11; ref. 11). The frequency of this naturally occurring broad-spectrum resistance to powdery mildew was 0.2–0.6% in total Ethiopian landrace material but in a particular locality was up to a level of 24% (refs 2, 12).

In contrast to fully resistant mutagen-induced *mlo*-null mutants³, *mlo*-11 plants allow the low-level growth of sporulating *Bgh* colonies (Fig. 1a). When homozygous *mlo*-11 resistant plants were self-pollinated (selfing), fully susceptible individuals were recovered with a frequency of about $(0.5-1) \times 10^{-4}$ (designated 'revertants'; Fig. 1a, Supplementary Table 1), indicating a possible meiotic instability of the *mlo*-11 allele. In contrast, no susceptible individual was found in about 125,000 progeny obtained after selfings of *mlo* resistant lines containing various mutation-induced lesions in *Mlo* (Supplementary Table 1). Extensive genetic analysis of the susceptible *mlo*-11 revertants indicated that either the *Mlo* susceptibility allele was restored or that susceptibility was the result of a heritable change in a tightly linked locus (Supplementary Table 2).

DNA sequencing of the *Mlo* coding region in *mlo*-11 resistant plants failed to detect differences from the *Mlo* wild-type sequence. However, genomic Southern blots probed with full-size *Mlo* complementary DNA (cDNA) detected expected fragment sizes of wild-type *Mlo* and additional strongly hybridizing fragments (Fig. 1b). Similarly, six of seven Ethiopian broad-spectrum powdery-mildew-resistant accessions of the Centre for Genetic Resources of The Netherlands (Supplementary Table 3), included in the haplotype analysis described below, showed a genomic Southern pattern identical to *mlo*-11 plants (not shown). The additional hybridizing signals were absent from both homozygous susceptible *mlo*-11 revertant progeny and susceptible *Mlo* wild-type control plants (Fig. 1b), indicating a causal link between the presence of these additional *Mlo*-homologous fragments in *mlo*-11 plants and resistance. Relative signal intensities suggested that the extra sequences were present in multiple copies. Polymerase chain reaction (PCR) analysis of *mlo*-11 genomic DNA showed that these were arranged as tandem repeat units, consisting of 1.1 kb of *Mlo* coding sequence (exon 1 to intron 5) flanked by 3.5-kb upstream sequences (Fig. 2a, b). Juxtaposed repeats were separated by a GT dinucleotide (Fig. 2b) not present in wild-type *Mlo*. Quantitative real-time PCR analysis revealed 9.4 ± 4.2 copies of the repeat unit in cultivar Ingrid BC *mlo*-11, 1.0 ± 0.3 copies in the tested homozygous susceptible revertant line, and 7.2 ± 4.3 copies in the tested homozygous resistant revertant sibling (all values relative to cultivar Ingrid *Mlo*, set as 1.0). The tandem repeat structure in *mlo*-11 is reminiscent of concatemers generated by the 'rolling-circle' DNA replication used by some viruses and transposons present in plants, for example the geminiviruses¹³ and *Helitron* transposons¹⁴. Chance use of a section of the *Mlo* gene by the rolling-circle DNA replication machinery offers a possible explanation for the presence of the *mlo*-11 repeat array.

We constructed a genomic cosmid library from *mlo*-11 plants and isolated four cosmid clones with the use of a *Mlo* 5'-terminal cDNA probe. DNA sequencing from the clones identified the 5' end of the repeat structure, consisting of a severely truncated repeat unit (Fig. 2d). Two low-copy loci located 5' of the repeat array were anchored to three of four *Mlo*-containing yeast artificial chromosome (YAC) clones, thereby unambiguously locating the *mlo*-11 repeat structure upstream of and adjacent to wild-type *Mlo* (Fig. 2e). None of the cosmid clones contained the 3' end of the repeat structure, and PCR amplification of this region from genomic DNA failed. However, the 3' end is likely to be located at least 1.8 kb upstream of the wild-type *Mlo* copy because identically sized fragments (representing wild-type *Mlo*) were detected by Southern

blot analysis in wild-type and *mlo-11* mutant plants by using diverse restriction enzymes (Fig. 1b, 2c). In addition, contiguous fibre-FISH fluorescent signals in chromosome sets derived from *Mlo* and *mlo-11* lines provided direct experimental evidence for an immediate physical linkage of the *mlo-11* repeat array and the *Mlo* wild-type copy (Fig. 1c).

We rescued two independent *mlo-11* revertants from resistant *mlo-11* accessions that flowered at different times. Both revertants retained their respective flowering time phenotypes, supporting the notion of an independent origin of the revertants (Supplementary Fig. 1). Although Southern analysis seemingly indicated a complete loss of the *mlo-11* repeat units in both revertants (Fig. 1b), examination by PCR revealed the presence of the truncated version of the 5'-terminal *mlo-11* repeat unit, thereby providing a molecular

footprint of ancestry (Supplementary Fig. 1b, c). Loss of the repeat units in the susceptible revertants might be explained by unequal crossover events of the 5'-most repeat unit with the wild-type *Mlo* gene copy during meiosis, a mechanism that would be compatible with the observed meiotic reversion frequency and the detected terminal repeat remnant in both revertants.

Although the repeat array in *mlo-11* plants is reminiscent of repeat-induced gene silencing, a phenomenon associated with transgene tandem repeat copies in eukaryotes¹⁵, we failed to detect significant CpG methylation or evidence for altered chromatin configuration at the *Mlo* locus in *mlo-11* plants (Supplementary Fig. 2). Heterozygous *Mlo/mlo-11* plants (hemizygous for the *mlo-11* repeat array) are susceptible to *Bgh* (Supplementary Table 2). This finding and transient expression analysis (Supplementary Table 4)

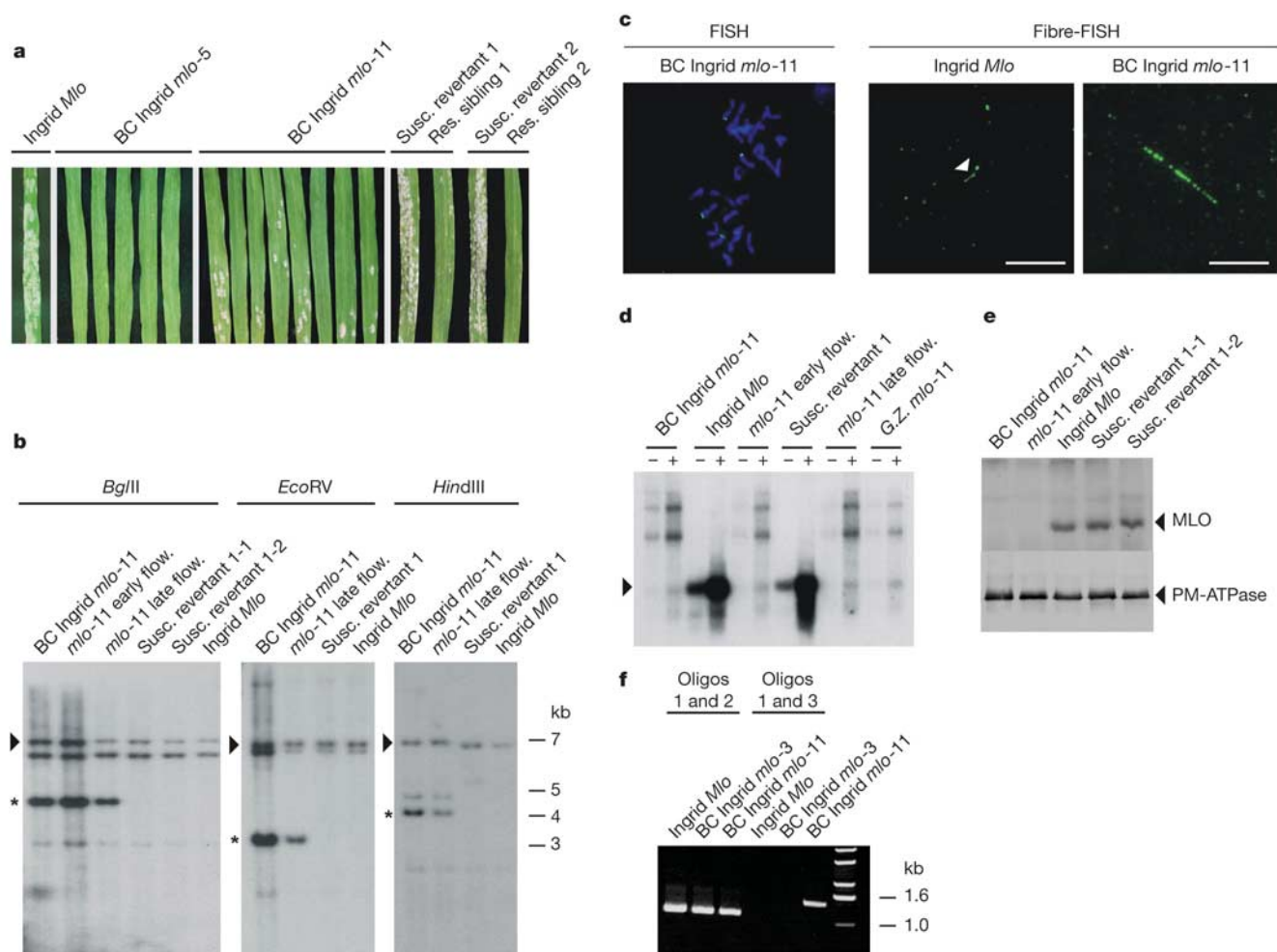


Figure 1 Phenotypic and molecular characterization of barley *Mlo* wild-type plants, resistant *mlo-11* mutants, susceptible *mlo-11* revertants and their resistant siblings. **a**, *Bgh* infection phenotypes of barley seedlings. Susc. revertant, homozygous (*Mlo/Mlo*) susceptible offspring of the heterozygous (*Mlo/mlo*) susceptible revertant; res. sibling, homozygous (*mlo/mlo*) resistant offspring of the heterozygous (*Mlo/mlo*) susceptible revertant. **b**, Southern blot probed with a full-size *Mlo* cDNA fragment. Arrows indicate expected wild-type bands; asterisks indicate *mlo-11*-specific bands. Additional bands probably result from cross-hybridization with the sequence-related *HvMlo2* homologue (GenBank accession number Z95496). **c**, FISH and fibre-FISH analysis (see Methods). Arrowhead, hybridization signal in Ingrid *Mlo*; scale bars, 10 μ m (corresponds to about 30 kb). **d**, Northern blot of poly(A)⁺ RNA from uninfected (–) or powdery-mildew-inoculated (+; samples taken 9 h after inoculation) barley seedlings probed with a full-size *Mlo* cDNA fragment. The arrowhead indicates the position of the ~2-kb *Mlo* wild-type

transcript. Note the aberrant high-molecular-mass transcripts in BC Ingrid *mlo-11*, *mlo-11* (late and early flowering) and G.Z. ('Grannenlose Zweizeilige') *mlo-11*. Susceptible revertant 1 is a homozygous (*Mlo/Mlo*) segregant of the susceptible heterozygous (*Mlo/mlo*) revertant obtained from *mlo-11* (late flowering) selfing. **e**, Western blot with protein extracts of enriched plasma membranes of uninfected barley seedlings probed with either MLO-specific (upper panel) or plasma membrane ATPase (PM-ATPase)-specific (lower panel) antisera as described previously⁴. **f**, RT-PCR analysis of RNA extracted from unchallenged barley seedlings with the use of the following oligonucleotides: 1, binds in *Mlo* cDNA downstream of repeat units, reverse primer; 2, binds in *Mlo* cDNA around the translational start site, forward primer; 3, binds at about –70 relative to the experimentally determined *Mlo* transcriptional start site³, forward primer.

reveals that the *mlo-11* repeat array does not interfere with the function of a wild-type *Mlo* copy *in trans*, thus excluding conventional post-transcriptional gene silencing as a potential interference mechanism. Inactivation of *Mlo* in *mlo-11* plants must therefore be fundamentally different from *trans*-acting epigenetic effects reported for the maize *b1* and *pericarp color1* loci, in which either tandem repeat units of 853 base pairs (bp) located 100 kb upstream of *b1* or the genetically unlinked *Unstable factor for orange1*, respectively, are required for epigenetic modification^{16,17}.

Constitutive low-level transcription of the *Mlo* gene is known to be upregulated about tenfold on challenge with a pathogen⁹. Wild-type *Mlo* transcripts were undetectable in poly(A)⁺ RNA isolated from unchallenged *mlo-11* resistant leaves (Fig. 1d). However, aberrant high-molecular-mass transcripts were observed that increased in abundance after challenge with a pathogen. At least part of these polyadenylated aberrant transcripts must originate from the supernumerary 5' regulatory sequences within the upstream *mlo-11* repeat units because we detected in reverse transcriptase PCR (RT-PCR) products of *mlo-11* RNA samples the *mlo-11*-specific GT dinucleotide that links adjacent repeat units at the DNA level (Fig. 1d, Supplementary Fig. 3). In addition, RT-PCR analysis provided direct evidence for transcriptional read-through of wild-type *Mlo* from upstream repeat units (Fig. 1f). Trace amounts of wild-type-sized transcripts were detectable in pathogen-challenged *mlo-11* leaves but these signals were still much weaker than those of wild-type transcripts in unchallenged susceptible *Mlo* plants (Fig. 1d). Transcript patterns in homozygous susceptible revertant progeny were undistinguishable from those in wild-type *Mlo* plants (Fig. 1d). Consistent with an essentially silent *Mlo* wild-type copy in *mlo-11* plants was our observation that western blot analysis failed to detect MLO protein in *mlo-11* mutants but revealed wild-type MLO concentrations in the homozygous susceptible revertants (Fig. 1e). These findings are strongly suggestive of a disruption of transcription machinery assembly by read-through into the wild-type *Mlo* copy from upstream repeat units in *mlo-11* plants¹⁸, a phenomenon not reported previously for a naturally occurring allele of a plant gene. The few powdery mildew colonies on *mlo-11* leaves (Fig. 1a) could indicate that the repeat-array-dependent transcriptional interference occasionally becomes 'leaky', an interpretation supported by the accumulation of very small amounts of wild-type-sized *Mlo* transcripts detected after challenge with a pathogen.

We performed haplotype analysis at the *Mlo* locus with a sample of 91 barley accessions comprising modern cultivated *H. vulgare*, undomesticated *H. spontaneum* ancestors, and cultivation intermediates represented by Ethiopian *H. vulgare* landraces. In a 25-kb interval including *Mlo* (ref. 19) we assayed eight sites exhibiting polymorphisms such as size variants of simple sequence repeats (SSRs) and the presence or absence of miniature transposable elements (MITEs). In addition, we monitored single-nucleotide polymorphisms (SNPs) as well as indels within the entire *Mlo* coding region (Supplementary Fig. 4 and Supplementary Table 5). In the cultivated accessions, three basic haplotypes (I–III) were observed that were distinct from one another across the entire interval (Fig. 3a). Representatives of each haplotype class were present both in early twentieth-century European varieties (I Binder, II Gull, III Hanna) and in more recently developed cultivars. A far greater number of haplotypes were identified within the *H. spontaneum* accessions (Fig. 3b). This pattern of diversity is consistent with a bottleneck associated with domestication when little of the genetic diversity of *H. spontaneum* was transferred to cultivated barley²⁰. All *mlo-11* resistant Ethiopian landrace accessions and modern European cultivars that carry the introgressed *mlo-11* allele contained near-identical haplotypes that grouped with class I haplotypes of wild-type *Mlo* cultivars (Fig. 3a and Supplementary Table 5). This is strongly indicative of a monophyletic

origin of *mlo-11* and indicates that the allele probably emerged recently within a common barley haplotype after domestication. The absence of SNPs between different *mlo-11* repeat units is consistent with the inferred recent emergence of the mutant allele (not shown).

Null mutations in *Mlo* have undesirable pleiotropic effects including accelerated leaf senescence and reduced grain yield^{2,9}, whereas the residual *Mlo* wild-type activity in *mlo-11* plants might reduce the severity of these effects. The southwestern part of Ethiopia, a region comprising highland areas with high rainfall, favours infections with powdery mildew¹². The adverse effects associated with *mlo*-null mutations and the capacity of polymorphic race-specific resistance loci to contain *Bgh* epidemics (for example, natural variation at *Mla*; ref. 21) might have prevented the elimination of the *Mlo* wild-type allele in undomesticated *H. spontaneum*. The presence of the broad-spectrum *mlo-11* resistance allele in primitive Ethiopian landraces might be advantageous because it would compensate for the inherent erosion of natural

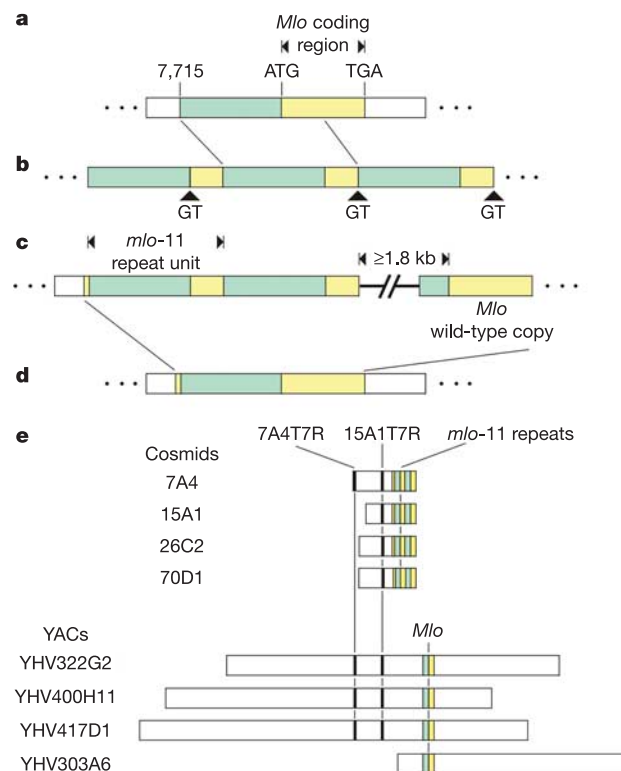


Figure 2 Schematic representation of the *Mlo* locus and *mlo-11* repeat array organization. **a**, The rectangle symbolizes the genomic locus at *Mlo*. *Mlo* coding sequence (11,211–14,049, numbering according to GenBank accession no. Y14573) is marked by a yellow box, the *Mlo* upstream sequence that is part of the repeat units (7,715–11,210) is indicated by a blue box. Flanking genomic regions are depicted in white. **b**, Three (of about five to ten) consecutive *mlo-11* repeat units. Colour coding of boxes is as described above. For simplicity, *Mlo* intron sequences are not indicated. GT indicates the *mlo-11*-specific GT dinucleotide separating each repeat unit. **c**, Linkage of terminal repeat units with flanking genomic regions. The most upstream repeat unit is severely truncated and connects to the authentic *Mlo* locus upstream region. The connection between the most downstream repeat unit and the *Mlo* wild-type gene copy is not resolved. **d**, Organization of the *Mlo* locus in susceptible revertants. The truncated upstream repeat unit identifies the revertant as a genuine *mlo-11* descendant. **e**, Relative arrangement of cosmid clones 7A4, 15A1, 26C2 and 70D1, cosmid-derived markers 7A4T7R and 15A1T7R, and *Mlo*-containing YAC clones YHV322G2, YHV400H11, YHV417D1 and YHV303A6 (ref. 26). Clones are not drawn to scale.

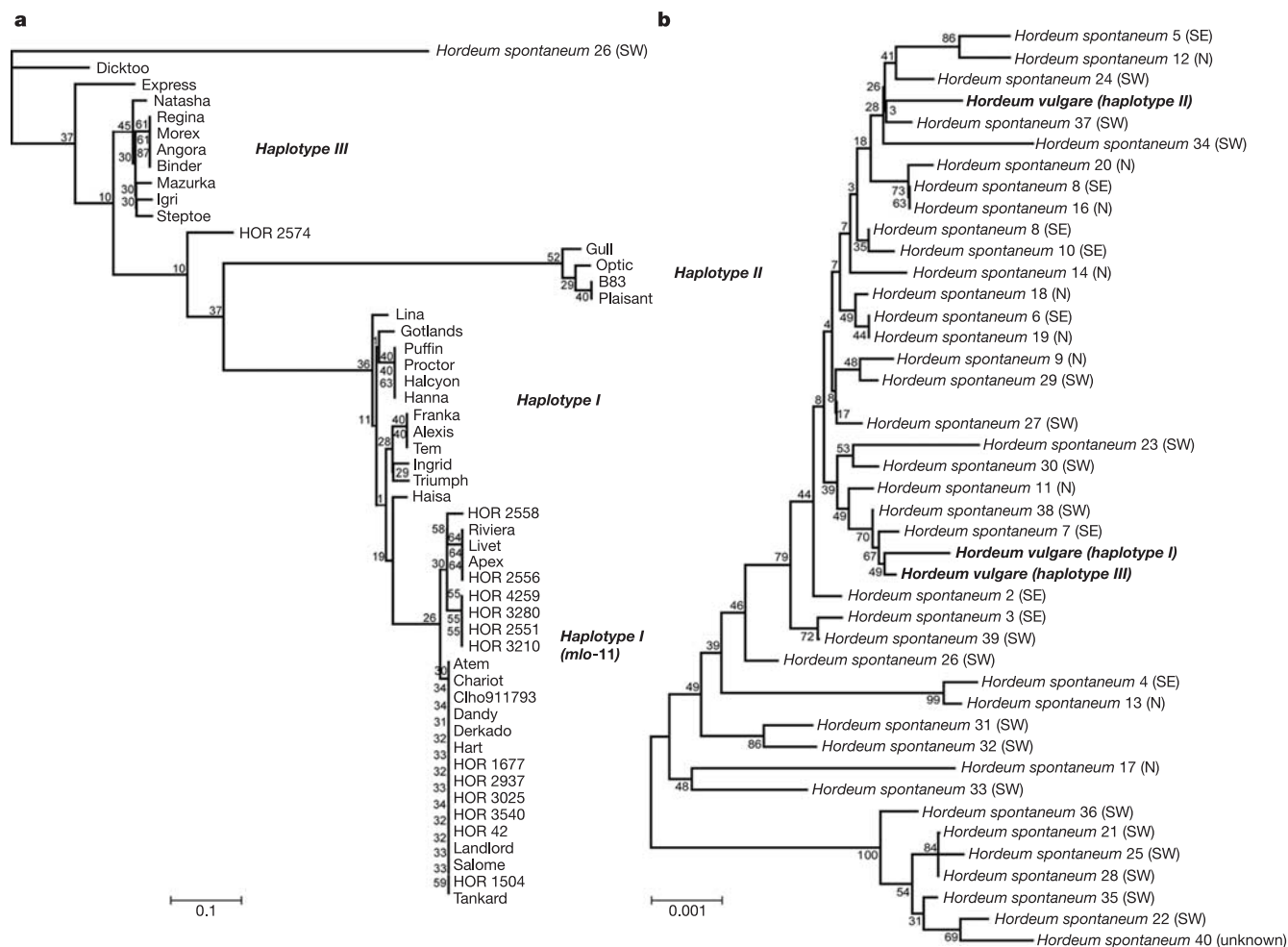


Figure 3 Phylogenetic relationship of barley germplasm based on haplotype analysis at *Mlo*. **a**, Phylogenetic tree based on genotypes of SSR and MITE markers in a 25-kb interval around *Mlo*. The scale bar indicates Nei's standard genetic distance D_s (ref. 27). **b**, Phylogenetic tree based on SNPs and indels in the *Mlo* coding region

(including introns). The scale bar indicates nucleotide substitutions per position. On both trees, the numbers above nodes indicate bootstrap support based on 10,000 replicates.

genetic variation at race-specific resistance loci upon barley domestication. It remains to be tested whether the occurrence and distribution of beneficial *mlo-11* resistance in the Ethiopian highlands resulted from conscious or unconscious selection by local farmers, which in effect might have created a cultivation-associated balanced polymorphism (*Mlo/mlo-11*) in barley landraces. □

Methods

Plant material

The germplasm used for haplotype analysis included modern European cultivars, some winter but mostly spring types, a subset of which contain introgressed *mlo-11* (ref. 22), some Ethiopian landrace accessions (*mlo-11* types) including probable donors of the resistance to cultivars² and a range of *Hordeum vulgare spontaneum* lines from Israel, Turkey and Iran²³. The Ethiopian landrace material was obtained from stock centres at the IPK Gatersleben (Germany), the Centre for Genetic Resources (Wageningen, The Netherlands) and the National Small Grains Collection (Aberdeen, Idaho, USA).

Barley genotypes Ingrid *Mlo*, backcross Ingrid (BC Ingrid) *mlo-11* and BC Ingrid *mlo-5* were kindly provided by J. MacKey (Uppsala, Sweden).

Haplotype analysis and generation of phylogenetic trees

In addition to a determination of the allelic variation at the *Mlo* locus by sequencing across all exons and introns (2,839 bp) the alleles were assayed at a range of putative polymorphic sites in the genic region. These included both SSRs and MITEs identified around the *Mlo* locus (Supplementary Fig. 4). SSR marker loci were amplified by PCR, respective products were resolved on polyacrylamide gels and the sizes of the products were estimated by reference to a SequaMark ladder (Research Genetics, Huntsville, Alabama, USA). The

MITE marker loci products were resolved on 1% agarose gels and the exact size differences were determined by sequencing of representative alleles. The polymorphisms found at these marker loci and within the gene sequence were combined to determine a haplotype of the *Mlo* region. The neighbour-joining tree shown in Fig. 3a was constructed with Populations version 1.2.28 (<http://www.cnrs-gif.fr/pge/bioinfo/populations/index.php>) using Nei's standard measure of distance D_s , and were visualized and drawn using Treeview version 1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>). *H. spontaneum* 26 (SW) served as outgroup for generation of the tree. The neighbour-joining tree shown in Fig. 3b was constructed using MEGA2 software (<http://www.megasoftware.net>) with Kimura 2 parameter model, both transitions and transversions included and the pairwise deletion option. Detailed data used for the generation of both phylogenetic trees can be found in Supplementary Table 5.

Generation of a *mlo-11* cosmid library

Genomic DNA of barley line *mlo-11* (early flowering) was partly digested with restriction enzyme *Sau3AI* and ligated into linearized (*Bam*HI) and dephosphorylated cosmid vector SuperCos1 (Stratagene, La Jolla, California, USA). Recombinant clones were found to carry inserts of about 30–42 kb and were sampled in 220 pools, each consisting of about 4,000 clones. Thus, the cosmid library represents about five barley genome equivalents (30 kb per clone × 220 pools × 4,000 clones ≈ 26.4 megabases (Mb); the barley genome size is about 5.3 Mb). The DNA sequence of presumptive (according to BLAST analysis) single/low-copy end fragments of cosmids 7A4 and 15A1 was exploited to derive oligonucleotides suitable for PCR amplification of the respective loci from genomic DNA.

Quantitative real-time PCR

Quantitative real-time PCR was used to estimate the number of *mlo-11* repeat units in genomic DNA of various barley genotypes. Oligonucleotide combinations that either specifically amplify *mlo-11* repeat DNA or a single-copy *Mlo* carboxy-terminal fragment

(serving as an internal standard) were used in independent reactions performed on an ABI PRISM 7700 Sequence Detection System (Applied Biosystems, Foster City, California, USA). Data were analysed by the comparative $\Delta\Delta C_T$ method (ABI PRISM 7700 User Bulletin) and represent means and standard deviations of five independent amplifications with two to four replicates each.

FISH and fibre-FISH analysis

Two mitotic metaphase chromosome sets of root tips derived from germinated seedlings of line BC Ingrid *mlo-11* were hybridized *in situ*²⁴ with a 4.5-kb *mlo-11* repeat unit fragment to demonstrate the specificity of this probe. For fibre-FISH analysis, nuclei were isolated from barley young leaves, fibres were prepared and stretched with the tilting method²⁵ and were subsequently hybridized with the 4.5-kb *mlo-11* repeat unit fragment.

Received 3 May; accepted 23 June 2004; doi:10.1038/nature02781.

1. Badr, A. *et al.* On the origin and domestication history of barley (*Hordeum vulgare*). *Mol. Biol. Evol.* **17**, 499–510 (2000).
2. Jørgensen, J. H. Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. *Euphytica* **63**, 141–152 (1992).
3. Büschges, R. *et al.* The barley *Mlo* gene: A novel control element of plant pathogen resistance. *Cell* **88**, 695–705 (1997).
4. Devoto, A. *et al.* Topology, subcellular localization, and sequence diversity of the *Mlo* family in plants. *J. Biol. Chem.* **274**, 34993–35004 (1999).
5. Devoto, A. *et al.* Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *J. Mol. Evol.* **56**, 77–88 (2003).
6. Panstruga, R. & Schulze-Lefert, P. Corruption of host seven-transmembrane proteins by pathogenic microbes: a common theme in animals and plants? *Microbes Infect.* **5**, 429–437 (2003).
7. Kim, M. C. *et al.* Calmodulin interacts with MLO protein to regulate defence against mildew in barley. *Nature* **416**, 447–450 (2002).
8. Collins, N. C. *et al.* SNARE-protein-mediated disease resistance at the plant cell wall. *Nature* **425**, 973–977 (2003).
9. Piffanelli, P. *et al.* The barley MLO modulator of defense and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiol.* **129**, 1076–1085 (2002).
10. Giessen, J. E., Hoffmann, W. & Schottenloher, R. Die Gersten Äthiopiens und Erythräas. *Z. Pflanzenzüchtung* **35**, 377–440 (1956).
11. Jørgensen, J. H. in *Barley Genetics III* (ed. Gaul, H.) 446–455 (Karl Thiemig, München, 1976).
12. Negassa, M. Geographic distribution and genotypic diversity of resistance to powdery mildew of barley in Ethiopia. *Hereditas* **102**, 113–121 (1985).
13. Gutierrez, C. Geminivirus DNA replication. *Cell. Mol. Life Sci.* **56**, 313–329 (1999).
14. Kapitonov, V. V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl Acad. Sci. USA* **98**, 8714–8719 (2001).
15. Henikoff, S. Conspiracy of silence among repeated transgenes. *BioEssays* **20**, 532–535 (1998).
16. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
17. Chopra, S. *et al.* The maize *Unstable factor for orange1* is a dominant epigenetic modifier of a tissue specifically silent allele of *pericarp color1*. *Genetics* **163**, 1135–1146 (2003).
18. Eszterhas, S. K., Bouhassira, E. E., Martin, D. I. K. & Fiering, S. Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol. Cell. Biol.* **22**, 469–479 (2002).
19. Panstruga, R., Büschges, R., Piffanelli, P. & Schulze-Lefert, P. A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. *Nucleic Acids Res.* **26**, 1056–1062 (1998).
20. Matus, I. A. & Hayes, P. M. Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. *Genome* **45**, 1095–1106 (2002).
21. Jørgensen, J. H. Genetics of powdery mildew resistance in barley. *Crit. Rev. Plant Sci.* **13**, 97–119 (1994).
22. Pakniyat, H. *et al.* AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt tolerance and associated ecogeography. *Genome* **40**, 332–341 (1997).
23. Thomas, W. T. B. *et al.* Identification of a QTL decreasing yield in barley linked to *Mlo* powdery mildew resistance. *Mol. Breed.* **4**, 381–393 (1998).
24. Benabdelmouna, A., Abirached-Darmency, M. & Darmency, H. Phylogenetic and genomic relationships in *Setaria italica* and its close relatives based on the molecular diversity and chromosomal organization of 5S and 18S–5.8S–25S rDNA genes. *Theor. Appl. Genet.* **103**, 668–677 (2001).
25. Fransz, P. F. *et al.* High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibers. *Plant J.* **9**, 421–430 (1996).
26. Simons, G. *et al.* AFLP-based fine mapping of the *Mlo* gene to a 30-kb DNA segment of the barley genome. *Genomics* **44**, 61–70 (1997).
27. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank B. Koop, C. Casais, I. Tierney and M. Macaulay for technical assistance; I. Somssich for experimental proposals; and N. Collins and M. Koornneef for suggestions on the manuscript. This work was supported by grants from the Gatsby Charitable Foundation to P.S.-L., from the Max-Planck Society to R.P., from Génoplatte to A.B., and from the Scottish Executive Environment and Rural Affairs Department and the European Commission to R.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.S.-L. (schlef@mpiz-koeln.mpg.de).

SNF-6 is an acetylcholine transporter interacting with the dystrophin complex in *Caenorhabditis elegans*

Hongkyun Kim¹, Matthew J. Rogers¹, Janet E. Richmond² & Steven L. McIntire¹

¹Ernest Gallo Clinic and Research Center, Programs in Neuroscience and Biomedical Sciences, Department of Neurology, University of California at San Francisco, 5858 Horton Street, Suite 200, Emeryville, California 94608, USA

²Department of Biological Sciences, University of Illinois at Chicago, Chicago, Illinois 60607, USA

Muscular dystrophies are among the most common human genetic diseases and are characterized by progressive muscle degeneration. Muscular dystrophies result from genetic defects in components of the dystrophin–glycoprotein complex (DGC), a multimeric complex found in the muscle cell plasma membrane¹. The DGC links the intracellular cytoskeleton to the extracellular matrix and is thought to be important for maintaining the mechanical integrity of muscles² and organizing signalling molecules³. The exact role of the DGC in the pathogenesis of disease has, however, remained uncertain⁴. Mutations in *Caenorhabditis elegans* DGC genes lead to specific defects in coordinated movement and can also cause muscle degeneration^{5–7}. Here we show that mutations in the gene *snf-6* result in phenotypes indistinguishable from those of the DGC mutants, and that *snf-6* encodes a novel acetylcholine/choline transporter. SNF-6 mediates the uptake of acetylcholine at neuromuscular junctions during periods of increased synaptic activity. SNF-6 also interacts with the DGC, and mutations in DGC genes cause a loss of SNF-6 at neuromuscular junctions. Improper clearing of acetylcholine and prolonged excitation of muscles might contribute to the pathogenesis of muscular dystrophies.

In a genetic screen⁸, we identified 12 mutants exhibiting a locomotory phenotype (defect in coordinated movement) indistinguishable from that of *dys-1* (a dystrophin homologue) mutants. These mutants consist of two alleles of *dys-1*, two alleles of *dyb-1* (a dystrobrevin homologue), three alleles of *dyc-1* (a dystrophin and CAPON-related gene) and five alleles of a previously uncharacterized gene (*eg28*, *eg114*, *eg115*, *eg121* and *eg137*). All of these mutants exhibit a locomotory phenotype distinct from that of other locomotion-defective mutants. During slow basal movements these mutants show essentially wild-type locomotory behaviours; however, when the mutants are forced to move forwards rapidly in response to mechanical stimulation, they exhibit exaggerated bending of the anterior body and head, and mild hypercontraction (Fig. 1; Supplementary videos 1, 2, 3, 5 and 6). Hypercontraction of body muscles is known to result from increased synaptic concentrations of acetylcholine at the neuromuscular junction (NMJ) in *C. elegans*. In fact, *dys-1* and *dyb-1* mutants are known to be mildly hypersensitive to aldicarb, an inhibitor of acetylcholinesterases that enhances muscle contraction, indicating elevated cholinergic synaptic transmission^{5,6}. *eg28* animals are also hypersensitive to aldicarb in comparison with wild-type animals (percentage paralysis of wild-type and *eg28* animals to 1 mM aldicarb: 3.3 ± 2.9 and 11.7 ± 2.9 at 50 min; 20 ± 5 and 35 ± 0 at 60 min; 38.3 ± 2.8 and 55 ± 8.7 at 70 min, $P < 0.02$). Furthermore, we were able to phenocopy the locomotory defect of these mutants by short-term exposure of wild-type animals to aldicarb (Fig. 1a). Together these observations indicate that the locomotory defect of these mutants might be due to elevated concentrations of acetylcholine at the NMJ.

To gain insight into the underlying abnormality of this class of

mutants, we cloned the gene disrupted by the *eg28* mutation by a combination of genetic mapping, RNA-mediated interference (RNAi) and transgenic rescue (Fig. 2a, b; see Supplementary Methods). The cloned gene (M01G5.5) encodes a putative transporter, previously designated *snf-6* solely on the basis of sequence similarity to the mammalian solute carrier 6 family (SLC6A) of sodium-dependent neurotransmitter transporters (GenBank accession number 661556). Alignment of the predicted amino acid sequence of *snf-6* reveals 30–40% identity to members of the SLC6A family. Mutation site analysis revealed that four of the five *snf-6* mutations (*eg28*, *eg115*, *eg121* and *eg137*) are nonsense mutations (Fig. 2c). *eg115* is an early nonsense mutation in the third transmembrane domain and is likely to be a null mutation. All of the *snf-6* mutants are phenotypically indistinguishable from one another.

To identify cells in which *snf-6* is expressed we generated transgenic animals expressing a *snf-6* green fluorescent protein (GFP) reporter construct. Strong expression in body wall, vulval and enteric muscles was observed (Fig. 2d). A few unidentified neuronal processes were also occasionally visible. To identify where SNF-6 was required, we attempted to rescue the *snf-6* locomotion phenotype by expressing *snf-6* in either neurons or muscles. Transgenic animals expressing *snf-6* under the pan-neuronal H20 promoter⁹ were not rescued. However, transgenic animals expressing *snf-6* under the control of the muscle-specific *myo-3* promoter were rescued for all of the phenotypes (Fig. 1a and Fig. 2b), indicating that *snf-6* might be required in muscle cells.

To identify the substrate of SNF-6 we performed uptake assays in a stably transfected HEK293 cell line expressing SNF-6. Because our analyses of *snf-6* mutants indicated possible defects in cholinergic neurotransmission, we tested whether SNF-6 transports acetylcholine. We observed a specific uptake of both acetylcholine and choline in a saturable manner ($K_m = 234.3 \pm 11.85 \mu\text{M}$ for acetylcholine and $K_m = 189 \pm 38.15 \mu\text{M}$ for choline) (Fig. 3a, b). Uptake was inhibited by the replacement of sodium with potassium, a signature of SLC6A-family transporters (Fig. 3c, d). We did not observe the uptake of other compounds, including several neurotransmitters (Supplementary Information). Consistent with this identified function of SNF-6 was the observation that GFP::SNF-6 expression was strongest in the distal end of muscle arms, where NMJs are located (Fig. 4b).

If the *snf-6* phenotype reflects defective clearance of acetylcholine, then changes in synaptic acetylcholine concentrations should affect the severity of the phenotype. Because mutations in *ace-1* or *ace-2* (genes encoding acetylcholinesterases) should increase the concentration of acetylcholine at NMJs, we hypothesized that double-mutant *snf-6;ace-1* and *snf-6;ace-2* animals would exhibit a more severe phenotype than *snf-6* single mutants if SNF-6 has a function in clearing acetylcholine at NMJs. The double mutants *snf-6;ace-1* and *snf-6;ace-2* had more exaggerated anterior body and head bends than *snf-6* single mutants (Supplementary Table 1, compare Supplementary videos 3 and 4). This enhancement of *snf-6* phenotypes was not observed with a *cho-1* mutation. *cho-1* encodes a presynaptic choline transporter, and *cho-1* mutants exhibit essentially wild-type locomotion (J. Rand, personal communication). These results indicate that SNF-6 might have a function in clearing acetylcholine under conditions causing elevated concentrations of acetylcholine at NMJs. This additive phenotype with *ace* mutants has been described with the other DGC mutants, *dyb-1* and *dys-1* (refs 10, 11).

For a direct analysis of the role of SNF-6 *in vivo* we quantified whole-cell voltage-clamped currents from NMJs of wild-type animals and *snf-6* mutants. Initial evoked synaptic currents recorded from muscles of *snf-6* mutants in response to ventral nerve cord stimulation were comparable to those recorded from wild-type animals (Fig. 5a). We suspected that the physiological consequences of a defect in clearance of acetylcholine might be magnified by

increased synaptic activity or repetitive stimulation. Indeed, absolute response amplitudes for the second to the fifth stimuli in a train of five stimuli delivered at a frequency of 20 Hz were significantly greater in *snf-6* than wild-type animals (Fig. 5b). Because an *ace-1* mutation enhances the *snf-6* phenotypes, we asked whether this enhancement is reflected in electrophysiological recordings of *snf-6;ace-1* mutants. The amplitude differences for the second to fifth responses were greater in *snf-6;ace-1* than in *ace-1* animals (Fig. 5c, d). The normalized difference between these two mutants was more marked than the difference between wild-type and *snf-6* mutants. Together these physiological results *in vivo* provide further confirmation of the identified role of SNF-6 in acetylcholine clearance when the rate of presynaptic acetylcholine release is increased. SNF-6 might also prevent spillover to adjacent synaptic sites.

Why do mutations in *snf-6* and the DGC genes lead to indistinguishable locomotory defects? We addressed this question by first constructing double mutants of *snf-6* with other DGC mutants and comparing the double-mutant phenotypes with those of the single mutants (Supplementary Table 1). The phenotype of *snf-6* was not enhanced or diminished by mutations in *dys-1*, *dyb-1* or *stn-1*.

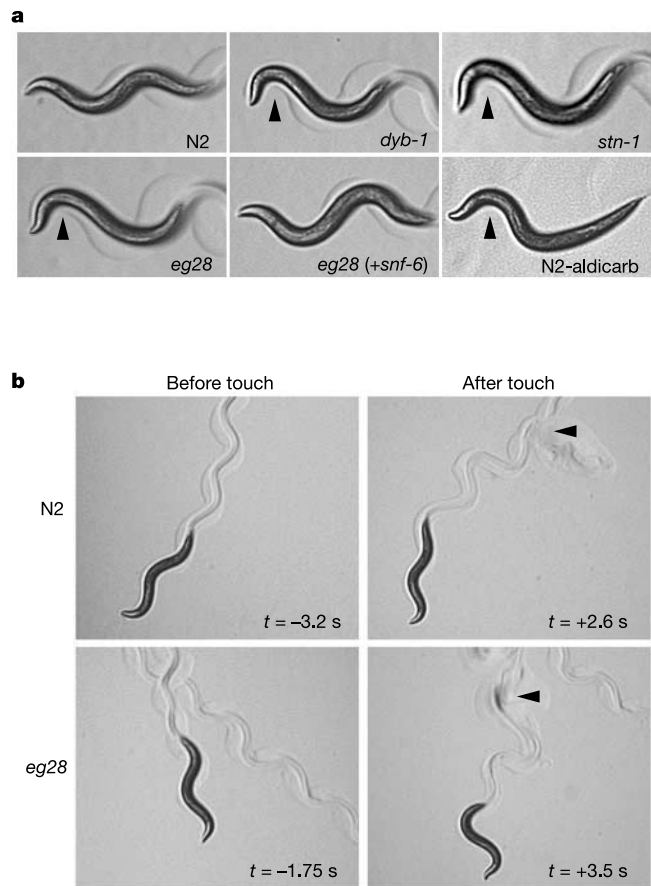


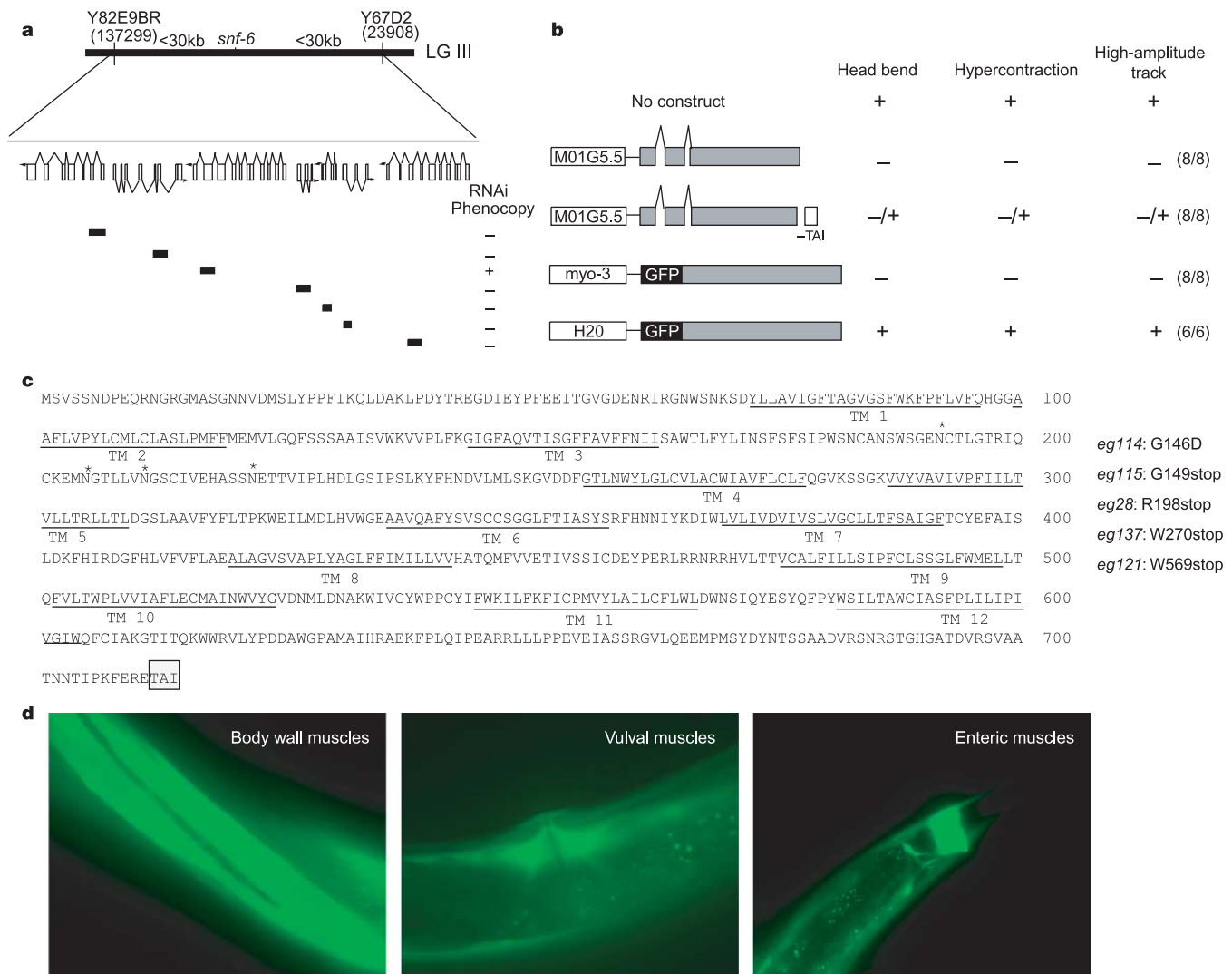
Figure 1 Phenotypic characterization of the DGC mutants and *eg28*. **a**, Still pictures were taken immediately after animals had been transferred (stimulated condition) to a new plate seeded with *Escherichia coli* (OP50). *snf-6* mutants rescued with a wild-type copy of the *snf-6* gene show wild-type locomotion, whereas DGC mutants and wild-type animals treated with 1 mM aldicarb show the characteristic exaggerated bending (arrowhead) of the anterior body and head. **b**, Still pictures of wild-type and *eg28* mutant animals from movies (Supplementary videos 5 and 6) showing locomotory changes before and after mechanical stimulation. Arrowhead, point of contact; *t*, time before or after stimulation.

These results are consistent with the possibility that *snf-6* functions in the same pathway as the other DGC components.

Interactions between the DGC and other effector molecules are often mediated by syntrophins. Syntrophins interact with dystrophin and α -dystrobrevin through a pleckstrin homology domain and with other effector molecules through a PDZ domain in mammals¹² and in *C. elegans*¹³. There is only one *C. elegans* syntrophin, *stn-1*, a β 1-syntrophin homologue⁷. The carboxy-terminal end of SNF-6, TIV, matches the consensus motif (T/S-X-I/V, where X is any amino acid) known to interact with the type I PDZ domain present in STN-1. We sought evidence for a direct physical interaction of SNF-6 with STN-1 in transiently transfected COS-7 cells. Reciprocal co-immunoprecipitation was observed with a combination of STN-1 and SNF-6, but not with STN-1 and SNF-6 Δ 3 (SNF-6 Δ 3 lacks the three C-terminal amino acids) (Fig. 4a). SNF-6 can therefore interact *in vitro* with STN-1 through the C-terminal PDZ interacting sequence. We tested the functional relevance of this interaction *in vivo* by attempting to rescue the *snf-6* phenotype by expressing SNF-6 Δ 3. These transgenic animals

showed only weak transient rescue or no rescue of the locomotory phenotype. Interaction of SNF-6 with STN-1 is likely to be important for wild-type locomotory behaviour (Fig. 2b).

If the DGC maintains SNF-6 at neuromuscular synapses, the distribution of SNF-6 should be altered in DGC mutants. We used an integrated P_{myo-3} GFP::SNF-6 reporter to compare the localization of SNF-6 in the *snf-6* mutant and in *stn-1*; *snf-6* and *dys-1*; *snf-6* double mutants. In 2-day-old adults we observed significant differences in GFP expression (Fig. 4b). Combined with a decrease in GFP expression, the fluorescence was more diffuse and often undetectable in the distal end of muscle arms in *stn-1* and *dys-1* mutant backgrounds (fluorescence intensity at end of muscle arms, arbitrary units: *snf-6*, 42.24 ± 3.33 , $n = 23$; *dys-1*; *snf-6*, 28.14 ± 1.61 , $n = 22$, $P < 0.002$; *stn-1*; *snf-6*, 23.91 ± 0.95 , $n = 24$, $P < 0.0001$). However, in early-stage adults there were more mild, variable changes in GFP expression. The DGC might therefore not be necessary to establish SNF-6 at postsynaptic specializations, but it is required to maintain or stabilize SNF-6 in the postsynaptic membrane. Examination of either GFP-tagged SGN-1 (sarcoglycan)



or STN-1 expression in muscles of wild-type and *snf-6* animals indicates that *snf-6* mutations do not cause any gross changes in the DGC (Supplementary Fig. S1).

In *C. elegans*, mutations in DGC genes result in progressive loss of the ability to move (partial or complete paralysis) and muscle degeneration in a sensitized *hlh-1* background. *hlh-1* is a homologue of myoD, a muscle-specific transcription factor. Similarly, mice lacking dystrophin (*mdx* mice) have a very mild myopathy, but mice lacking myoD and dystrophin have a severe myopathy¹⁴. If the loss of SNF-6 function at NMJs contributes to muscle degeneration in these double mutants, mutations in both *snf-6* and *hlh-1* should also lead to muscle degeneration and a progressive paralysis. *hlh-1;snf-6* double mutants exhibit the same progressive paralysis as is observed

in *dyb-1;hlh-1* double mutants¹¹ (Fig. 4c). Analysis of body muscles of *hlh-1;snf-6* double mutants (fourth-day adults) (Fig. 4d) also revealed an equivalent degree of muscle degeneration to that found in *dys-1;hlh-1* and *dyb-1;hlh-1* double mutants^{11,15} (percentage of animals scored as positive: *snf-6*, 0%, *n* = 26; *hlh-1*, 0%, *n* = 41; *hlh-1;snf-6*, 17%, *n* = 74) (Fig. 4d). The degeneration in *hlh-1;snf-6* was reduced significantly (7% positive, *n* = 105) by RNAi of *egl-19* (L-type calcium channel in muscle), indicating that the muscle degeneration might be dependent on muscle excitation. RNAi of *egl-19* causes a similar reduction of muscle degeneration in *dys-1;hlh-1* (ref. 16). Furthermore, chronic treatment of *hlh-1* mutants with a low dose of aldicarb led to muscle degeneration, whereas treatment of wild-type animals did not (*hlh-1*, 20%, *n* = 126; N2, 0%, *n* = 52) (Fig. 4d). Together these findings indicate a role for altered transport of acetylcholine in the muscle degeneration that results from a loss of DGC function in *C. elegans*.

We have identified an acetylcholine transporter that functions in neurotransmission during periods of elevated synaptic activity.

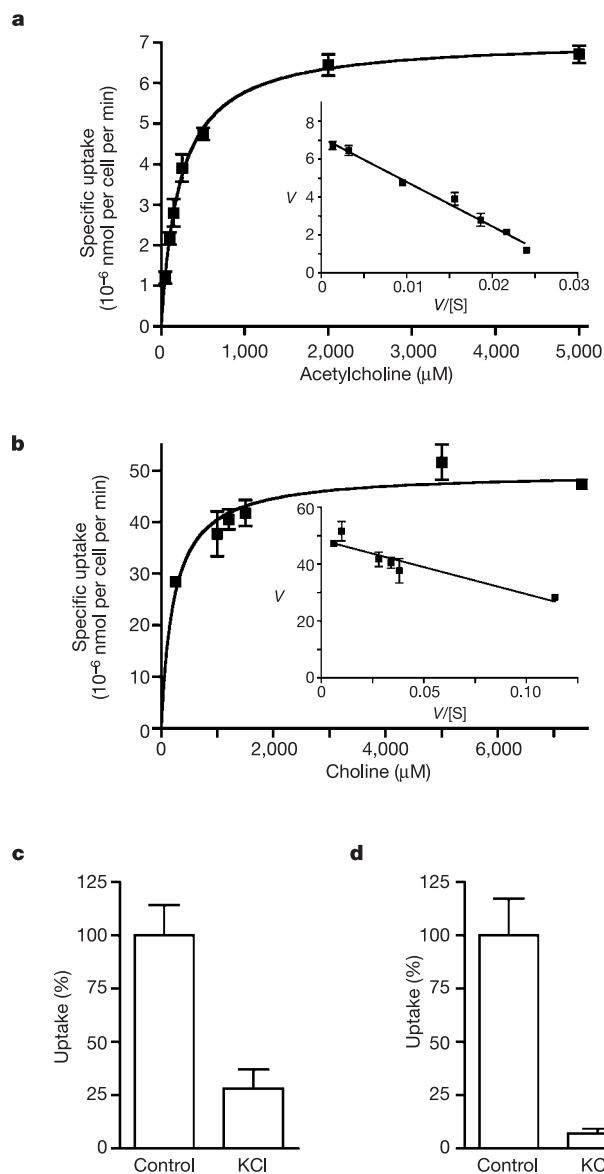


Figure 3 Biochemical characterization of SNF-6 by uptake assays. Results are means $\pm$ s.e.m. from triplicate trials with a stably transfected cell line expressing SNF-6 after subtracting the non-specific uptake from triplicate trials of a control cell line. **a, b**, Saturation curve of acetylcholine (**a**) and choline (**b**). Insets, Eadie-Hofstee transformation of saturation data. *V*, velocity; *[S]*, substrate concentration. **c, d**, Sodium-dependent uptake of acetylcholine (**c**) and choline (**d**). The cells were incubated for 15 min with 10 nM acetylcholine or 20 nM choline in assay buffer containing either NaCl (control) or KCl. Values are expressed as a percentage of control.

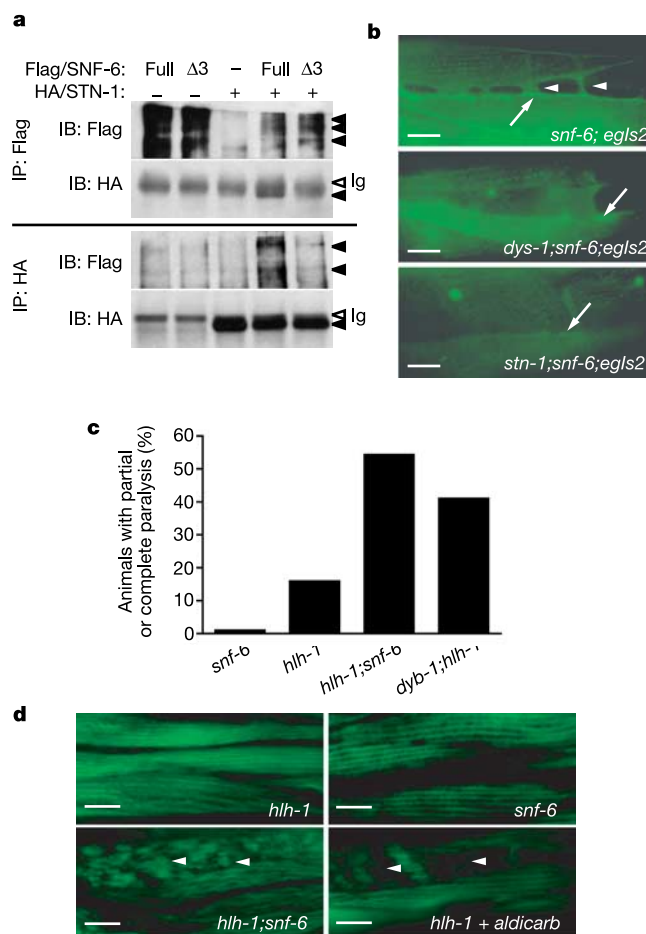


Figure 4 SNF-6 is important for muscle integrity. **a**, Interaction of SNF-6 with STN-1. Full, full-length SNF-6; Δ3, SNF-6 lacking three C-terminal amino acids; IP, immunoprecipitation; IB, immunoblot; arrowhead, signals for Flag or HA epitopes; open arrowhead, Ig heavy chain. **b**, Localization of GFP::SNF-6 in mutants (48 h after L4). *egl-2*, *egl-2*^{P_{myo-3}::GFP::snf-6}, *pRF4(rol-6d)*; arrowheads, muscle arms; arrow, ends of muscle arms where NMJs occur. Fluorescent intensity comparison in *snf-6;egl-2* (NMJs, 42.24 $\pm$ 3.30; membrane area, 26.04 $\pm$ 1.01, *P* < 0.05). **c**, Four-day-old adults failing to move at least three body lengths after stimulation were scored as paralysed (*snf-6*, *n* = 64; *hlh-1*, *n* = 44; *hlh-1;snf-6*, *n* = 46; *dyb-1;hlh-1*, *n* = 56). **d**, Muscle organization in mutants. Arrowheads, degenerated muscle fibres. Scale bars, 10 μm.

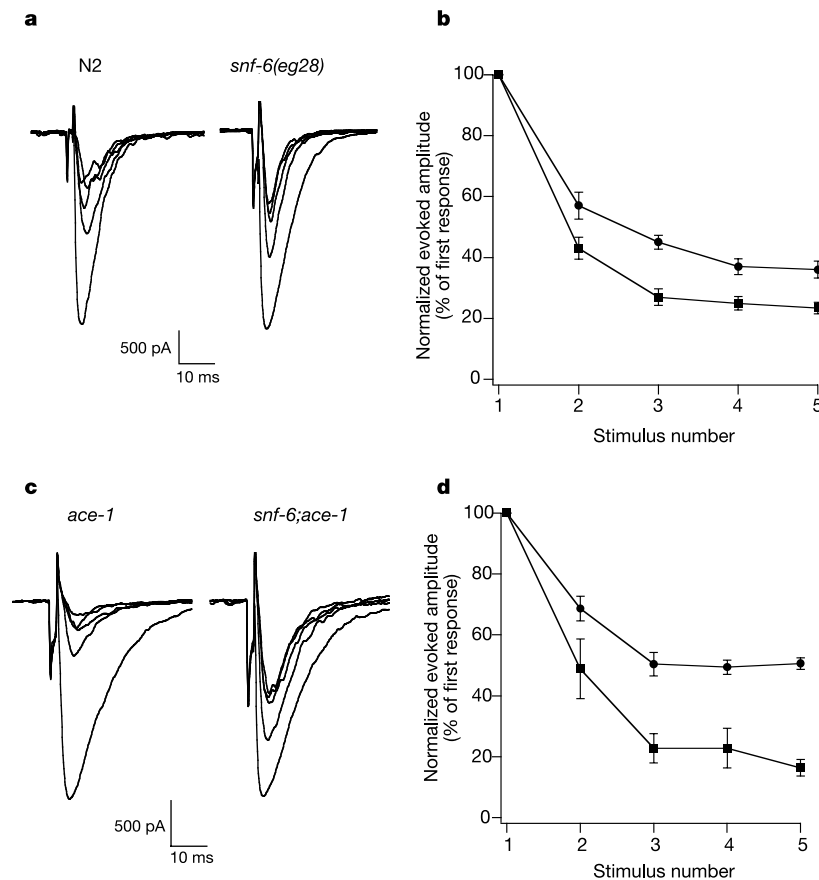


Figure 5 *snf-6* mutants exhibit enhanced evoked synaptic responses. **a, c**, Recordings from muscles of wild-type, *snf-6*, *ace-1* and *snf-6;ace-1* animals in response to five depolarizing stimuli delivered to the ventral nerve cord (20 Hz). **b, d**, The amplitude of each evoked response in the train was normalized to the initial response amplitude for

each recording: comparison of normalized data for fifth evoked response in **b** (wild-type (squares), $23.5 \pm 1.9\%$, $n = 17$; *snf-6* (circles), $35.5 \pm 2.8\%$, $n = 8$, $P < 0.01$) and **d** (*ace-1* (squares), $16.4 \pm 2.7\%$, $n = 5$; *snf-6;ace-1* (circles), $50.6 \pm 1.86\%$, $n = 5$, $P < 0.01$). All statistically derived values are given as means $\pm$ s.e.m.

Neurotransmitter transporters have previously been identified for most of the classical neurotransmitters. A transporter of acetylcholine was not thought to be necessary because of the high efficiency of acetylcholinesterase and the presynaptic reuptake of choline by the high-affinity choline transporter, CHO-1. Nevertheless, certain earlier studies in vertebrates have indicated a possible uptake of acetylcholine into muscle at cholinergic synapses^{17,18}.

Loss of SNF-6-mediated acetylcholine transport provides an explanation for the defects in cholinergic synaptic function, periodic muscle hyperexcitation, and calcium-dependent muscle degeneration observed in the DGC mutants. In mammalian systems, acute hyperexcitation of muscles through inhibition of acetylcholinesterases for 1–2 h is known to result in an endplate myopathy¹⁹ that can be prevented by the inactivation of acetylcholine receptors or a decrease in calcium concentration²⁰. Abnormal calcium accumulation has variably been observed in dystrophic mammalian muscles and has been postulated to have a function in the degenerative process in Duchenne muscular dystrophy (DMD)²¹. Interaction of the *C. elegans* SNF-6 acetylcholine transporter with the DGC is mediated through STN-1, a β 1-syntrophin homologue. In mammals β 1-syntrophin is expressed only at the NMJs of fast-twitching type IIB muscles²², which are engaged under conditions required to produce large force within a brief period. Interestingly, fast-twitching type IIB fibres are preferentially affected in human patients with DMD²³ and *mdx* mice²⁴. A central problem in defining the pathogenesis of muscular dystrophy has been the

lack of understanding of how disruption of the DGC leads to muscle degeneration. Our results indicate that in *C. elegans*, loss of SNF-6-mediated acetylcholine uptake and prolonged muscle excitation contribute to the muscle degeneration resulting from disruption of the DGC. □

Methods

Strains and germ-line transformation

All the strains were maintained at 20 °C using standard methods²⁵. We used the following strains: wild-type N2 Bristol isolate, LS292 *dys-1(cx18)* I, LS505 *dyb-1(cx35)* I, LS721 *stn-1(ok292)* I, GG202 *ace-2(g72)* I, PD4605 *hlh-1(cc561)* II, BZ28 *snf-6(eg28)* III, FX373 *cho-1(tm373)* IV, PR1000 *ace-1(p1000)* X, BZ692 *snf-6(eg28)* III; *eglIs2[P_{myo-3}::GFP::snf-6, pRF4(rol-6d)]*.

Germ-line transformation was performed by injecting test DNA into the gonads of *snf-6(eg28)* or wild-type N2 animals at a concentration of 10 ng μ l⁻¹, unless stated otherwise, together with markers *ofm-1::GFP*²⁶ or pRF4 at 30 ng μ l⁻¹.

Uptake assay

FLAG epitope-tagged *snf-6* complementary DNA was subcloned into pLNCX2 (BD Bioscience). This construct was transfected into a retroviral packaging cell line PT67, which was used for obtaining viral particles. Infectious viral particles were used to generate stable HEK293 cell lines expressing SNF-6. A clone was chosen on the basis of the level of expression in western analysis with anti-Flag antibodies. For uptake assays, cells expressing control vector or SNF-6 were plated on a six-well plate (10⁶ cells per well) and cultured overnight. The cells were washed twice with an assay buffer (10 mM HEPES pH 7.4, 4.7 mM KCl, 2.2 mM CaCl₂, 1.2 mM MgSO₄, 1.8 mg ml⁻¹ glucose), then incubated with assay buffer containing the indicated amount of [³H]acetylcholine (85 Ci mmol⁻¹ or 76 mCi mmol⁻¹) or [³H]choline (86 Ci mmol⁻¹) over 15 min at 28 °C. Uptake reactions were terminated by washing cells five times with ice-cold assay buffer. The cells were then solubilized with 1% SDS, and the accumulated radioactivity was measured with a scintillation counter (Beckman).

Electrophysiology

Electrophysiological methods were performed as described previously^{27,28}. In brief, animals were immobilized with a cyanoacrylic glue and a lateral incision was made to expose the ventral nerve cord and body wall muscles. Muscle recordings were made in the whole-cell voltage-clamp configuration (holding potential -60 mV) at room temperature (19°C) with an EPC-10-2 patch-clamp amplifier (HEKA, Lambrecht, Germany) and digitized at 2.9 kHz by means of an ITC-16 interface (Instrutech, Great Neck, New York, USA). Data were acquired by Pulse software (HEKA). The bath solution contained 150 mM NaCl, 5 mM KCl, 5 mM CaCl_2 , 1 mM MgCl_2 , 10 mM glucose and 15 mM HEPES pH 7.35 ; ~ 330 mOsm. The pipette solution contained 120 mM KCl, 20 mM KOH, 4 mM MgCl_2 , 5 mM (N -tris(hydroxymethyl)methyl-2-aminoethane-sulphonic acid), 0.25 mM CaCl_2 , 4 mM NaATP, 36 mM sucrose, 5 mM EGTA pH 7.2 ; ~ 315 mOsm. Subsequent analysis and graphing were performed using Pulsefit (HEKA), Mini Analysis (Synaptosoft) and Igor Pro (Wavemetrics, Lake Oswego, Oregon, USA).

Immunoprecipitation and western blotting

snf-6 and *stm-1* cDNAs, tagged with Flag and haemagglutinin (HA) epitopes, respectively, were subcloned to a mammalian expression vector. An equal amount of each construct was transfected into COS-7 cells with FUGENE 6 (Roche). The transfected cells were lysed with lysis buffer (50 mM Tris-HCl pH 7.4 , 1% Nonidet P40, 0.25% sodium deoxycholate, 10% glycerol, 150 mM NaCl and a mixture of protease inhibitors) and the resulting lysates were used for immunoprecipitation with anti-Flag (Roche) or anti-HA (Sigma) antibodies. Immune complexes were resolved by 8% SDS-polyacrylamide-gel electrophoresis and then transferred to poly(vinylidene difluoride) membrane. Each blotted membrane was cut into two different parts for western blotting with anti-HA and anti-FLAG antibodies.

Staining and microscopy

For actin fibre staining, 4-day-old adult animals grown at 15°C were fixed with 4% paraformaldehyde for 16 h and treated with collagenase ($2,000$ U ml $^{-1}$)²⁹, then stained with Alexa Fluor 488 phalloidin (Molecular Probes). Animals were scored as having degenerated muscles if we observed disorganized muscle fibres in at least two muscle quadrants (40 cells). Aldicarb treatment was performed on plates from the L1 larval stage at a concentration ($10\text{ }\mu\text{M}$) that does not cause complete paralysis.

To observe fluorescence from transgenic animals expressing GFP, animals were mounted on a 4% agar pad containing sodium azide. In the cases that required quantification of the levels of fluorescence in different genetic backgrounds, a SPOT CCD camera was used to capture images for a fixed time interval, and pixel intensity at the NMJs was determined for each sample with NIH image 1.63. At least 30 animals were randomly chosen for quantification of fluorescence levels in each line. The percentage of animals showing detectable expression did not vary significantly between lines (data not shown).

Received 11 May; accepted 2 July 2004; doi:10.1038/nature02798.

- Durbeej, M. & Campbell, K. P. Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models. *Curr. Opin. Genet. Dev.* **12**, 349–361 (2002).
- McArdle, A., Edwards, R. H. & Jackson, M. J. Time course of changes in plasma membrane permeability in the dystrophin-deficient *mdx* mouse. *Muscle Nerve* **17**, 1378–1384 (1994).
- Grady, R. M. *et al.* Maturation and maintenance of the neuromuscular synapse: genetic evidence for roles of the dystrophin-glycoprotein complex. *Neuron* **25**, 279–293 (2000).
- Burton, E. A. & Davies, K. E. in *Pathogenesis of Neurodegenerative Disorders* (ed. Mattson, M. P.) 239–284 (Humana Press, Totowa, New Jersey, 2000).
- Bessou, C., Giuglia, J. B., Franks, C. J., Holden-Dye, L. & Segalat, L. Mutations in the *Caenorhabditis elegans* dystrophin-like gene *dys-1* lead to hyperactivity and suggest a link with cholinergic transmission. *Neurogenetics* **2**, 61–72 (1998).
- Gieseler, K., Bessou, C. & Segalat, L. Dystrobrevin- and dystrophin-like mutants display similar phenotypes in the nematode *Caenorhabditis elegans*. *Neurogenetics* **2**, 87–90 (1999).
- Grisoni, K., Martin, E., Gieseler, K., Mariol, M. C. & Segalat, L. Genetic evidence for a dystrophin-glycoprotein complex (DGC) in *Caenorhabditis elegans*. *Gene* **294**, 77–86 (2002).
- Davies, A. G. *et al.* A central role of the BK Potassium channel in behavioral responses to ethanol in *C. elegans*. *Cell* **115**, 655–666 (2003).
- Shioi, G. *et al.* Mutations affecting nerve attachment of *Caenorhabditis elegans*. *Genetics* **157**, 1611–1622 (2001).
- Giuglia, J., Gieseler, K., Arpagaus, M. & Segalat, L. Mutations in the dystrophin-like *dys-1* gene of *Caenorhabditis elegans* result in reduced acetylcholinesterase activity. *FEBS Lett.* **463**, 270–272 (1999).
- Gieseler, K. *et al.* Molecular, genetic and physiological characterisation of dystrobrevin-like (*dys-1*) mutants of *Caenorhabditis elegans*. *J. Mol. Biol.* **307**, 107–117 (2001).
- Froehner, S. C., Adams, M. E., Peters, M. F. & Gee, S. H. Syntrophins: modular adapter proteins at the neuromuscular junction and the sarcolemma. *Soc. Gen. Physiol. Ser.* **52**, 197–207 (1997).
- Gieseler, K., Abdel-Dayem, M. & Segalat, L. *In vitro* interactions of *Caenorhabditis elegans* dystrophin with dystrobrevin and syntrophin. *FEBS Lett.* **461**, 59–62 (1999).
- Megency, L. A., Kablar, B., Garrett, K., Anderson, J. E. & Rudnicki, M. A. MyoD is required for myogenic stem cell function in adult skeletal muscle. *Genes Dev.* **10**, 1173–1183 (1996).
- Gieseler, K., Grisoni, K. & Segalat, L. Genetic suppression of phenotypes arising from mutations in dystrophin-related genes in *Caenorhabditis elegans*. *Curr. Biol.* **10**, 1092–1097 (2000).
- Mariol, M. C. & Segalat, L. Muscular degeneration in the absence of dystrophin is a calcium-dependent process. *Curr. Biol.* **11**, 1691–1694 (2001).
- Potter, L. T. Synthesis, storage and release of [^{14}C]acetylcholine in isolated rat diaphragm muscles. *J. Physiol. (Lond.)* **206**, 145–166 (1970).
- Saelens, J. K. & Stoll, W. R. Radiochemical determination of choline and acetylcholine flux from isolated tissue. *J. Pharmacol. Exp. Ther.* **147**, 336–342 (1965).
- Salpeter, M. M., Kasprzak, H., Feng, H. & Fertuck, H. Endplates after esterase inactivation *in vivo*: correlation between esterase concentration, functional response and fine structure. *J. Neurocytol.* **8**, 95–115 (1979).

- Leonard, J. P. & Salpeter, M. M. Agonist-induced myopathy at the neuromuscular junction is mediated by calcium. *J. Cell Biol.* **82**, 811–819 (1979).
- Jackson, M. J., Jones, D. A. & Edwards, R. H. Measurements of calcium and other elements in muscle biopsy samples from patients with Duchenne muscular dystrophy. *Clin. Chim. Acta* **147**, 215–221 (1985).
- Kramarcy, N. R. & Sealock, R. Syntrophin isoforms at the neuromuscular junction: developmental time course and differential localization. *Mol. Cell. Neurosci.* **15**, 262–274 (2000).
- Webster, C., Silberstein, L., Hays, A. P. & Blau, H. M. Fast muscle fibers are preferentially affected in Duchenne muscular dystrophy. *Cell* **52**, 503–513 (1988).
- Karpati, G., Carpenter, S. & Prescott, S. Small-caliber skeletal muscle fibers do not suffer necrosis in *mdx* mouse dystrophy. *Muscle Nerve* **11**, 795–803 (1988).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Miyabayashi, T., Palfreyman, M. T., Sluder, A. E., Slack, F. & Sengupta, P. Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. *Dev. Biol.* **215**, 314–331 (1999).
- Richmond, J. E. & Jorgensen, E. M. One GABA and two acetylcholine receptors function at the *C. elegans* neuromuscular junction. *Nature Neurosci.* **2**, 791–797 (1999).
- Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2**, 959–964 (1999).
- Waterston, R. H., Hirsh, D. & Lane, T. R. Dominant mutations affecting muscle structure in *Caenorhabditis elegans* that map near the actin gene cluster. *J. Mol. Biol.* **180**, 473–496 (1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. S. Kim, C. Yu and R. Ho for technical help, J. Rand and C. Bargmann for personal communication and helpful discussions, and all members of the McIntire laboratory for discussions and comments on the manuscript. Some strains were provided by the National Bioresource Project (Japan), the *C. elegans* Gene Knockout Consortium and the *Caenorhabditis* Genetics Center. This work was supported by funds provided by the State of California for medical research on alcohol and substance abuse through the University of California, San Francisco, by a grant to S.L.M. from the Department of the Army, by a grant to J.E.R. from NIH and by a development grant to H.K. from the Muscular Dystrophy Association.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.L.M. (e-mail: slm@itsa.ucsf.edu).

Coupling of agonist binding to channel gating in an ACh-binding protein linked to an ion channel

Cecilia Bouzat¹, Fernanda Gumilar¹, Guillermo Spitzmaul¹, Hai-Long Wang², Diego Rayes¹, Scott B. Hansen³, Palmer Taylor³ & Steven M. Sine²

¹Instituto de Investigaciones Bioquímicas, UNS-CONICET, Bahía Blanca 8000, Argentina

²Receptor Biology Laboratory, Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, Minnesota 55905, USA

³Department of Pharmacology, University of California, San Diego, La Jolla, California 92093-0636, USA

Neurotransmitter receptors from the Cys-loop superfamily couple the binding of agonist to the opening of an intrinsic ion pore in the final step in rapid synaptic transmission. Although atomic resolution structural data have recently emerged for individual binding¹ and pore domains², how they are linked into a functional unit remains unknown. Here we identify structural requirements for functionally coupling the two domains by combining acetylcholine (ACh)-binding protein, whose structure was determined at atomic resolution¹, with the pore domain from the serotonin type-3A (5-HT_{3A}) receptor. Only when amino-acid sequences of three loops in ACh-binding protein are changed to their 5-HT_{3A} counterparts does ACh bind with low affinity characteristic of activatable receptors, and trigger opening of the ion pore. Thus functional coupling requires structural compatibility at the interface of the binding

and pore domains. Structural modelling reveals a network of interacting loops between binding and pore domains that mediates this allosteric coupling process.

Pentameric ligand-gated ion channels mediate rapid synaptic transmission throughout the nervous system, and include receptors activated by ACh, γ -aminobutyric acid, glycine and 5-HT (refs 3–5). Their vital role in converting chemical recognition into an electrical impulse makes these receptors prime loci for learning, memory and

disease processes, as well as targets for clinically relevant drugs. Activation of synaptic receptors comprises three elementary steps: agonist recognition, coupling to the ion pore and opening of the pore. Each step has a physical counterpart, and recent X-ray crystallographic and electron microscopic data cast considerable light on the atomic structures of the binding¹ and pore counterparts². However, the structural basis for functionally coupling binding and pore domains remains a topic of intense investigation^{2,6,7}.

To investigate the structural basis for functionally coupling binding and pore domains, we generated a chimaeric receptor composed of the ACh binding domain from acetylcholine binding protein (AChBP) and the pore domain from the 5-HT_{3A} receptor (Fig. 1a). AChBP was chosen because its atomic structure is known¹, and it presumably evolved without the constraint of functional coupling to an ionic pore, whereas the 5-HT_{3A} pore enables expression of homo-pentameric receptors in mammalian cells^{8,9}. After transfection of the chimaeric complementary DNA into Bosc 23 cells, robust quantities of the receptor chimaera are expressed on the surface of the cells, as measured by α -bungarotoxin binding, and ACh binds to the chimaera with micro-molar affinity (Table 1). However, reliable ACh-induced currents could not be detected using the patch clamp. We therefore reasoned that although binding and pore domains were correctly folded, enabling expression on the cell surface, the interface between the two domains was not compatible, preventing inter-domain coupling.

From the known atomic coordinates of AChBP (ref. 1) and the pore domain of the *Torpedo* acetylcholine receptor², we generated a homology model of our chimaeric receptor (Fig. 1b, c). Each subunit contains an AChBP module composed of ten β -strands and interconnecting loops, and a pore domain containing four α -helices. The model shows three of the interconnecting loops from the binding domain project into or near the pore domain, and the loop connecting M2 and M3 helices of the pore domain projects into the binding domain. We hypothesized that interplay between these four loops is crucial for functional coupling.

To test this hypothesis, we mutated residues in AChBP to their 5-HT_{3A} counterparts in the three loops that face the pore, thus generating two new chimaeras designated C3 and C3L (Table 1 and Fig. 1a); these chimaeras are identical except that C3L contains three additional residues of the 5-HT_{3A} sequence substituted in the β 8– β 9 loop. Both chimaeras express robust quantities of cell surface receptors in Bosc 23 cells, and show a 20- to 40-fold increase in apparent dissociation constant (K_d) for ACh binding (Table 1). Expression amounts and apparent K_d for ACh are distinct from those of the AChBP/5-HT_{3A} chimaera, but are similar to those of a

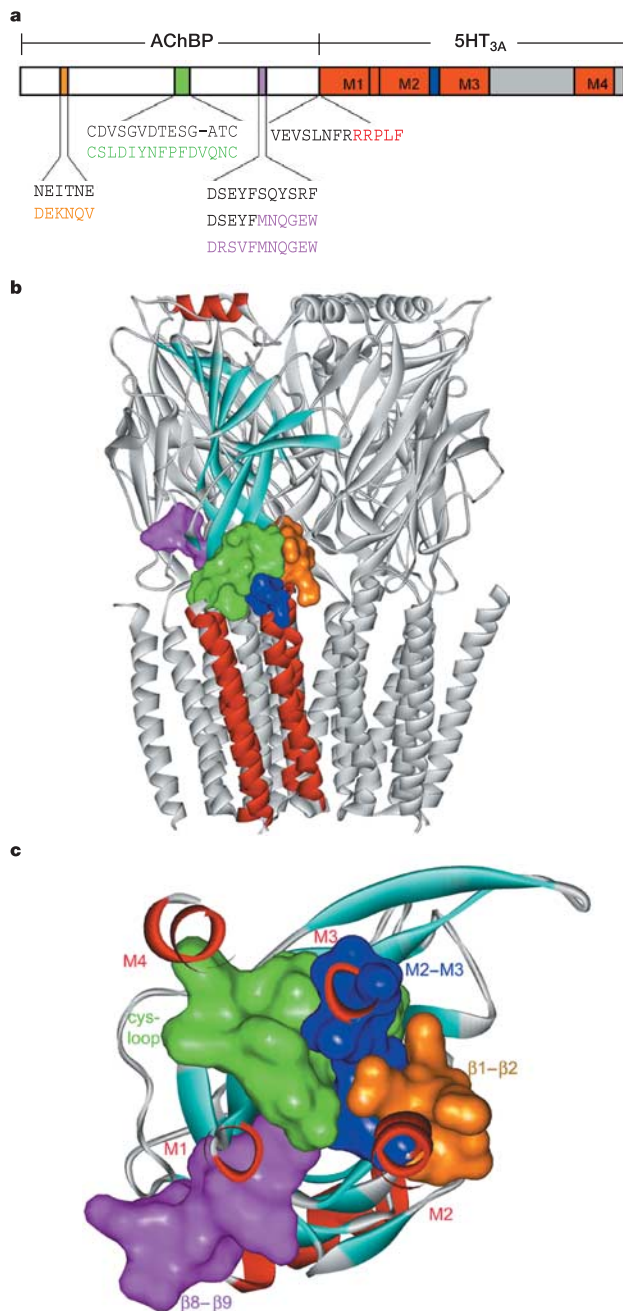


Figure 1 Design and structural model of receptor chimaeras. **a**, Schematic diagram of a subunit composed of AChBP and 5-HT_{3A} sequences. The amino terminus contains AChBP (white bars and black font) and 5-HT_{3A} sequences (loop β 1– β 2, orange; Cys-loop, green; β 8– β 9 loop, magenta). The 5-HT_{3A} pore domain follows, with transmembrane helices M1–M4 (red) and the M2–M3 linker (blue). Red font indicates the start of the M1 domain. **b**, Homology model of the chimaeric receptor (C3, Table 1) with key regions of one subunit colour-coded as in **a**. **c**, View of the coupling zone from the pore showing the network of binding and pore domain loops.

Chimaera	β 1– β 2	Cys-loop	β 8– β 9	Expression (%)	ACh K_d (μ M)
AChBP/5-HT _{3A}	AChBP	AChBP	AChBP	100	7.1 $\pm$ 2.6 (10)
C1 β 1–2	5-HT _{3A}	AChBP	AChBP	120 $\pm$ 20 (6)	7.6 $\pm$ 0.4 (2)
C1 Cys	AChBP	5-HT _{3A}	AChBP	ND (8)	NM
C1 β 8–9	AChBP	AChBP	5-HT _{3A}	6.5 $\pm$ 1.5 (4)	NM
C1 β 8–9L	AChBP	AChBP	5-HT _{3A} (L)	3.6 $\pm$ 1.9 (4)	NM
C2 β 1–2, Cys	5-HT _{3A}	5-HT _{3A}	AChBP	ND (8)	NM
C2 β 1–2, β 8–9	5-HT _{3A}	AChBP	5-HT _{3A}	160 $\pm$ 14 (4)	15.8 $\pm$ 1.5 (1*)
C2 β 1–2, β 8–9L	5-HT _{3A}	AChBP	5-HT _{3A} (L)	11.0 $\pm$ 5.0 (4)	NM
C2 Cys, β 8–9	AChBP	5-HT _{3A}	5-HT _{3A}	8.1 $\pm$ 1.9 (4)	NM
C2 Cys, β 8–9L	AChBP	5-HT _{3A}	5-HT _{3A} (L)	4.5 $\pm$ 2.0 (4)	NM
C3	5-HT _{3A}	5-HT _{3A}	5-HT _{3A}	140 $\pm$ 26 (6)	155 $\pm$ 52 (4)
C3L	5-HT _{3A}	5-HT _{3A}	5-HT _{3A} (L)	190 $\pm$ 30 (6)	280 $\pm$ 60 (3)
alpha7/5-HT _{3A}	alpha7	alpha7	alpha7	350 $\pm$ 32 (4)	138 $\pm$ 10 (1*)

C1, C2 or C3 indicate chimaeras with single, double or triple loop substitutions of 5-HT_{3A} sequence. L indicates that β 8– β 9 linker has three additional residues of 5-HT_{3A} sequence (Fig. 1a). Expression was measured by [¹²⁵I]- α -bungarotoxin binding to intact Bosc 23 cells, and normalized relative to that of the AChBP/5-HT_{3A} chimaera. ACh binding was measured by competition against [¹²⁵I]- α -bungarotoxin binding, yielding the apparent K_d from fitting the Hill equation to the data. Values are means $\pm$ s.d. and the number of determinations is in parenthesis. Asterisks indicate where standard error of the fitted parameter is given. ND, not detectable; NM, not measurable owing to low or undetectable expression.

chimaera in which the ligand binding domain of the homomeric $\alpha 7$ receptor is joined to the 5-HT_{3A} pore domain (Table 1). On the other hand, when residues in the three loops are substituted in AChBP alone, the resulting soluble protein, C3Ltrunc, shows no change in ACh affinity (AChBP, $K_d = 4.6 \mu\text{M}$; C3Ltrunc, $K_d = 5.8 \mu\text{M}$), showing that residue substitution in the three loops does not directly affect ACh binding, but rather alters affinity by coupling between binding and pore domains. Moreover, ACh evokes inward currents when rapidly applied to cells expressing C3 or C3L, as well as our reference receptor $\alpha 7/5\text{-HT}_{3A}$ (Fig. 2a). Thus low affinity for ACh and functional coupling between binding and pore domains require structural compatibility at the inter-domain interface (Fig. 1c).

To resolve ACh-evoked currents at the level of single channels, we noted that although unitary currents through 5-HT_{3A} receptors are too small to be resolved, this was overcome by neutralizing positively charged residues in the M3–M4 linker on the cytoplasmic face of the membrane¹⁰. We therefore mutated these arginines to uncharged or negatively charged residues in the chimaeras AChBP/5-HT_{3A}, C3, C3L and $\alpha 7/5\text{-HT}_{3A}$ (see Methods), and looked for ACh-evoked single-channel currents using the cell-attached configuration of the patch clamp. ACh elicits clear-cut single-channel activity in membrane patches from cells expressing

conductance-enhanced C3, C3L and $\alpha 7/5\text{-HT}_{3A}$ chimaeras (Fig. 2b); no unitary currents could be detected for the construct derived from AChBP/5-HT_{3A}, despite recording from many patches from cells in which protein expression was confirmed by the presence of green fluorescent protein. Dwell time histograms of the single-channel currents show that the $\alpha 7/5\text{-HT}_{3A}$ and C3 chimaeras open and close with distinct kinetics (Fig. 2c), as might be expected from their different ligand binding domains.

Although the modular composition of our chimaeric receptors accounts for distinct kinetics of pore opening and closing, flow of ions through the pore of each chimaera should be similar⁸. To test this hypothesis, we measured amplitudes of ACh-evoked single-channel currents over a range of membrane potentials. Highly similar slope conductances are observed for conductance-enhanced versions of $\alpha 7/5\text{-HT}_{3A}$, C3 and C3L, and moderate inward rectification is observed for each, as expected from their common pore domain (Fig. 3).

Our results show functional coupling when amino-acid sequences in all three loops from AChBP are changed to their

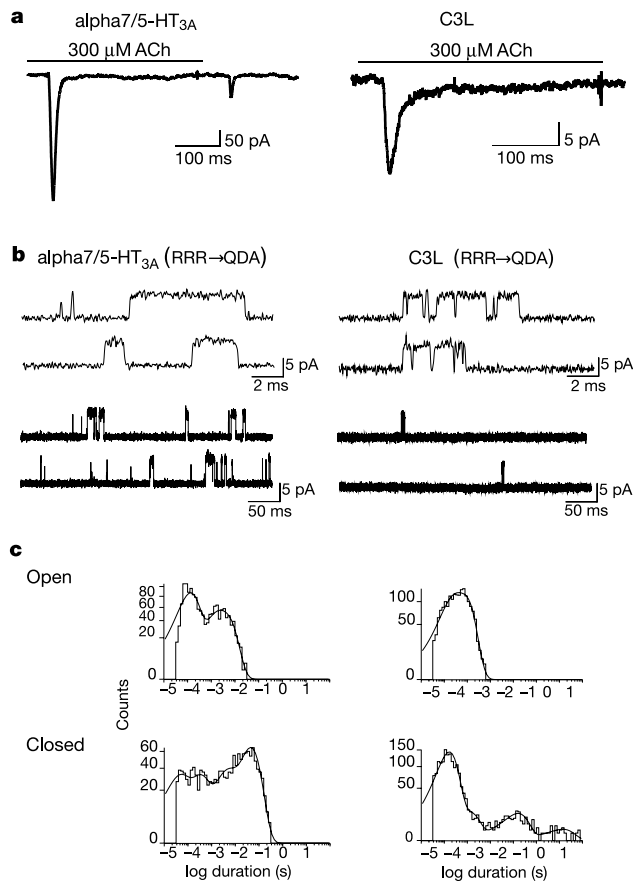


Figure 2 ACh-activated currents through chimaeric receptors. **a**, Whole-cell currents through $\alpha 7/5\text{-HT}_{3A}$ or C3L chimaeras (300 μM ACh). Each trace represents the average of 4 to 8 ACh applications. Decay time constants are 12 ms ($\alpha 7/5\text{-HT}_{3A}$) and 15 ms (C3L). Holding potential, -50 mV . **b**, Single-channel currents through $\alpha 7/5\text{-HT}_{3A}$ and C3L chimaeras from cell-attached patches (1 mM ACh). Chimaeras contain the triple mutation (RRR $\rightarrow$ QDA) to increase conductance¹⁰. Currents are displayed at 9 kHz bandwidth with channel openings as upward deflections. Membrane potential, -70 mV . **c**, Open and closed duration histograms from currents in **b**.

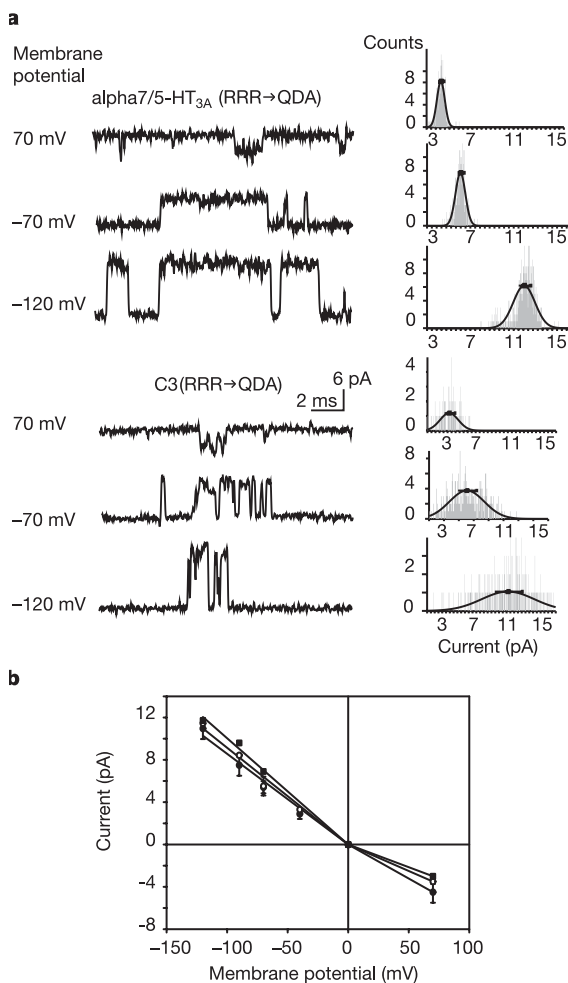


Figure 3 ACh-activated single-channel currents through chimaeric receptors at different membrane potentials. **a**, Unitary currents activated by 1 mM ACh (left) and corresponding amplitude histograms (right) are shown for each chimaeric receptor at the indicated membrane potentials. Bars indicate means $\pm$ s.d. **b**, Current-voltage relationships for $\alpha 7/5\text{-HT}_{3A}$ (open circles), C3 (closed circles) and C3L (open squares) containing the triple mutation (RRR $\rightarrow$ QDA). Each point represents the mean $\pm$ s.d. from three different patches. Data were fitted by linear regression. Inward and outward slope conductances are: $\alpha 7/5\text{-HT}_{3A}$ (RRR $\rightarrow$ QDA), 99 and 57 pS; C3 (RRR $\rightarrow$ QDA), 91 and 50 pS; and C3L (RRR $\rightarrow$ QDA), 98 and 43 pS.

5-HT_{3A} counterparts. To determine whether changes in all three loops are required, we generated chimaeras in which only one or two loops in the AChBP module are changed to 5-HT_{3A} sequence, and measured cell surface expression and apparent affinity for ACh (Table 1). Single loop substitutions of either the signature Cys-loop or the β 8– β 9 linker yielded undetectable or very low expression on the cell surface, suggesting impaired protein folding due to incompatibility within the binding domain or between binding and pore domains. For chimaeras showing undetectable expression on the cell surface, α -bungarotoxin binding to cells made permeable with saponin was also not detectable (data not shown), suggesting impaired folding rather than defective receptor trafficking. Substitution of the β 1– β 2 linker alone afforded expression of cell-surface receptors similar to AChBP/5-HT_{3A}, but affinity for ACh was unaltered, unlike the functionally coupled C3 and C3L chimaeras, which showed large increases in apparent K_d (Table 1). Conversion of this single-loop chimaera to the conductance-enhanced version, as described for C3 and C3L, afforded robust expression on the cell surface, but no ACh-evoked unitary currents could be detected from cells in which protein expression was confirmed. Four of five of the combinations of double loop substitutions yielded either undetectable or very low expression on the cell surface. The exception to this is the double-loop construct containing 5-HT_{3A} sequences in the β 1– β 2 and β 8– β 9 linkers, which shows robust expression. However, this construct shows only a twofold increase in apparent K_d for ACh, distinct from the large increases observed for the functionally coupled C3 and C3L chimaeras (Table 1), and no ACh-evoked single channel currents could be detected from cells expressing the conductance-enhanced version. Thus all three loops from AChBP have to be changed to their 5-HT_{3A} counterparts to produce high expression on the cell surface, low affinity for ACh and functional coupling between binding and pore domains.

The collective findings show a synergistic interplay between three loops from the binding domain and the M2–M3 loop from the pore domain, which is required for efficient expression, shift to low affinity for ACh and opening of the ion pore. Interplay between these loops couples the conformational change at the agonist binding site to the ion pore, and conversely, couples the functional state of the pore to the binding site. Thus the four loops within the coupling zone mediate a bi-directional allosteric interaction between the agonist binding site and distal ion pore.

The chimaeras with compatible inter-domain loops, C3 and C3L, activate in response to ACh and desensitize in its continued presence (Fig. 2), and both active and desensitized states are expected to bind the agonist with high affinity³. However, efficient activation requires low affinity of the resting compared with active and desensitized states¹¹, as well as predominance of the resting state in the absence of ACh. Our steady-state ACh binding measurements reflect the dynamic equilibrium between resting, active and desensitized states. The chimaera with an unmodified AChBP module shows high affinity compared to C3 and C3L chimaeras, but does not show ACh-evoked currents, suggesting its dynamic equilibrium is biased towards the high affinity desensitized state³. This bias could arise from significant accumulation of the desensitized state in the absence of ACh, and complete accumulation in its presence. On the other hand, the C3 and C3L chimaeras show low affinity for ACh, suggesting that although they desensitize, their dynamic equilibrium is sufficiently biased towards the low affinity resting state to enable activation by ACh.

Our structural model places the four inter-domain loops proximal to each other, and suggests interactions between particular loops within the subunit as well as between subunits (Fig. 1b, c). Although precise inter-residue contacts require further experimental testing, several general conclusions can be drawn about inter-loop interactions. First, the interplay between binding domain loops occurs through direct or indirect contact among all three loops within the subunit. While the β 1– β 2 and Cys-loops make

direct contact, the β 8– β 9 loop contacts the Cys-loop indirectly through the intervening strand β 10, and projects its side chains into the hydrophobic core of the subunit, impinging upon core side chains from the β 1– β 2, Cys- and M2–M3 loops (Fig. 1c and Supplementary Information). Furthermore, the β 1– β 2 loop from each subunit contacts the β 8– β 9 loop from the adjacent subunit (Fig. 1b). Second, the interplay between loops from the binding and pore domains has striking implications for functional coupling. The binding domain loops β 1– β 2 and Cys-loop straddle opposite sides of the M2–M3 linker from the pore domain (Fig. 1c), suggesting a twisting motion during coupling that in turn twists the pore-lining M2 domain¹² to allow flow of cations. This twisting motion is probably initiated by motions in parts of the binding domain distal to this region¹³, including perturbations due to ACh in the binding pocket and shearing of contacts at subunit interfaces. Additional parts of the pore domain may also contribute to coupling, such as the beginning of the M1 domain, which contacts the Cys-loop. Third, structural compatibility among the four inter-domain loops is likely to be a general requirement for allosteric coupling in rapidly-gated synaptic receptors. Finally, AChBP has the capacity for conformational change upon ACh binding that triggers opening of an ion pore, showing its utility beyond that of a structural surrogate for the ligand binding domain of the Cys-loop receptor superfamily. □

Methods

Construction of chimaeric AChBP/5-HT_{3A} subunits

The cDNA encoding AChBP (ref. 14) was subcloned into the cytomegalovirus-based expression vector pRBG4 (ref. 15), and contained the alpha7 signal sequence¹⁶. The pore domain from the rat 5-HT_{3A} receptor was joined to AChBP by bridging a *Bsi*WI restriction site at the 3' end of AChBP and a mutagenically-installed *Kas*I site just 5' from nucleotides encoding the M1 domain of the rat 5-HT_{3A} receptor, using synthetic double-stranded oligonucleotides¹⁷. The amino-acid sequence at the M1 junction is shown in Fig. 1a. Conversion of AChBP to 5-HT_{3A} sequence was performed by successive cycles of the QuikChange Site-Directed Mutagenesis kit (Stratagene). To increase unitary conductance for single-channel recordings, we mutated three arginine residues responsible for the low conductance of the 5-HT_{3A} receptors¹⁰ as follows: R432Q, R436D and R440A. All mutant and chimaeric subunits were confirmed by restriction enzyme analysis and sequencing.

Expression of chimaeric receptors

Bosc 23 cells¹⁸ were transfected with the subunit cDNAs using calcium phosphate precipitation, as described previously¹⁷. A plasmid encoding green fluorescent protein (pGreen lantern) was included to identify transfected cells under fluorescence optics. Cells were studied 2–3 days after transfection.

ACh binding measurements

The number of [¹²⁵I] α -bungarotoxin sites on the cell surface of transfected cells and ACh competition against the initial rate of [¹²⁵I] α -bungarotoxin were determined as previously described¹⁹. ACh binding to water-soluble AChBP and C3Ltrunc was determined as described¹⁶. Nonspecific binding was determined in the presence of 500 μ M D-tubocurarine. The Hill equation was fitted to the competition measurements to yield the apparent dissociation constant, K_d and the Hill coefficient. To measure binding of α -bungarotoxin to intracellular complexes, cells were first made permeable with 0.5% saponin (w/v) (ref. 20).

Single-channel recording

Recordings were obtained in the cell-attached configuration at 20 °C. Extracellular and pipette solutions contained 142 mM KCl, 5.4 mM NaCl, 0.2 mM CaCl₂ and 10 mM HEPES (pH 7.4). Single-channel currents were recorded using an Axopatch 200B patch-clamp amplifier (Axon Instruments, Inc.), digitized at 5 μ s intervals with the PCI-6111E interface (National Instruments), recorded to hard disc using the program Acquire (Bruxton Corporation) and detected by the half-amplitude threshold criterion using the program TAC (Bruxton Corporation) at a final bandwidth of 10 kHz (ref. 21). Open and closed time histograms were plotted using a logarithmic abscissa and a square-root ordinate²² and fitted to the sum of exponentials by maximum likelihood.

Macroscopic current recording

Macroscopic currents were recorded in the whole-cell configuration²³ using a perfusion system that allowed rapid (0.1–1 ms) solution exchange^{24,25}. The pipette solution contained 140 mM KCl, 5 mM EGTA, 5 mM MgCl₂ and 10 mM HEPES, pH 7.3. Extracellular solution contained 150 mM NaCl, 5.6 mM KCl, 0.2 mM CaCl₂ and 10 mM HEPES, pH 7.4. A series of 200–400 ms applications of ACh was applied to each cell. Macroscopic currents were filtered at 5 kHz, digitized at 20 kHz and analysed using IgorPro software (WaveMetrics Inc.). The following exponential function was fitted to

average currents from four to eight ACh applications: $I(t) = I_0 \exp(-t/\tau_d) + I_\infty$ where I_0 and I_∞ are the peak and the steady state currents, respectively, and τ_d is the time constant of current decay.

Homology modelling

A homology model of our receptor chimera (Fig. 1b, c) was generated using version 6.0 of the program MODELER²⁶, together with spatial restraints provided by AChBP¹ and *Torpedo* acetylcholine receptor pore domain² structures, as described²⁷. Briefly, we aligned by homology the sequence of the chimera C3 (Fig. 1a) with that of AChBP fused to the pore domain of the *Torpedo* receptor; the A subunit from AChBP was joined to the pore domain of the α_1 subunit, B to gamma, C to α_2 , D to delta and E to beta. To maintain complementarity between subunits at their interfaces, all five subunits were modelled simultaneously. We used the 'patch' command in MODELER to constrain coordinates of cysteines 128 and 142, which form a disulphide bond in each subunit. Among several options in MODELER, we selected the 'refine1 mode', which generates the highest level of refinement using conjugate gradients coupled with simulated annealing and molecular dynamics. Modelling included all polar hydrogens to allow for main chain hydrogen bonding, but omitted non-polar hydrogens.

Received 17 May; accepted 14 June 2004; doi:10.1038/nature02753.

1. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
2. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423**, 949–955 (2003).
3. Sine, S. M. The nicotinic receptor ligand binding domain. *J. Neurobiol.* **3**, 431–446 (2002).
4. Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nature Rev. Neurosci.* **3**, 102–114 (2002).
5. Changeux, J. P. & Edelstein, S. J. Allosteric receptors after 30 years. *Neuron* **21**, 959–980 (1998).
6. Grosman, C., Salamone, F., Sine, S. M. & Auerbach, A. The extracellular linker of muscle acetylcholine receptors is a gating control element. *J. Gen. Physiol.* **116**, 327–339 (2000).
7. Kash, T., Jenkins, A., Kelly, J., Trudell, J. & Harrison, N. L. Coupling of agonist binding to channel gating in the GABA_A receptor. *Nature* **421**, 272–275 (2003).
8. Eisele, J. L. *et al.* Chimeric nicotinic-serotonergic receptor combines distinct ligand binding and channel specificities. *Nature* **366**, 479–483 (1993).
9. Quiram, P. & Sine, S. M. Identification of residues in the neuronal $\alpha 7$ receptor that confer selectivity for conotoxin ImI. *J. Biol. Chem.* **273**, 11001–11006 (1998).
10. Kelley, S., Dunlop, J., Kirkness, E., Lambert, J. & Peters, J. A. A cytoplasmic region determines single-channel conductance in 5-HT₃ receptors. *Nature* **424**, 321–324 (2003).
11. Jackson, M. B. Perfection of a synaptic receptor: kinetics and energetics of the acetylcholine receptor. *Proc. Natl Acad. Sci. USA* **86**, 2199–2203 (1989).
12. Unwin, N. Acetylcholine receptor channel imaged in the open state. *Nature* **373**, 37–43 (1995).
13. Unwin, N., Miyazawa, A., Li, J. & Fujiyoshi, Y. Activation of nicotinic acetylcholine receptor involves a switch in conformation of the alpha subunits. *J. Mol. Biol.* **319**, 1165–1176 (2002).
14. Hansen, S. B. *et al.* Tryptophan fluorescence reveals conformational changes in the acetylcholine binding protein. *J. Biol. Chem.* **277**, 41299–41302 (2002).
15. Lee, B. S., Gunn, R. B. & Kopito, R. R. Functional differences among nonerythroid anion exchangers expressed in a transfected human cell line. *J. Biol. Chem.* **266**, 11448–11454 (1991).
16. Gao, F. *et al.* Curariform antagonists bind in different orientations to acetylcholine binding protein. *J. Biol. Chem.* **278**, 23020–23026 (2003).
17. Sine, S. M. Molecular dissection of subunit interfaces in the acetylcholine receptor: identification of residues that determine curare selectivity. *Proc. Natl Acad. Sci. USA* **90**, 9436–9440 (1993).
18. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
19. Sine, S. M. & Taylor, P. Functional consequences of agonist-mediated state transitions in the cholinergic receptor. *J. Biol. Chem.* **254**, 3315–3325 (1979).
20. Prince, R. J. & Sine, S. M. Molecular dissection of subunit interfaces in the acetylcholine receptor: identification of residues that determine agonist selectivity. *J. Biol. Chem.* **271**, 25770–25777 (1996).
21. Wang, H.-L. *et al.* Acetylcholine receptor M3 domain: stereochemical and volume contributions to channel gating. *Nature Neurosci.* **2**, 226–233 (1999).
22. Sigworth, F. & Sine, S. M. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* **52**, 1047–1054 (1987).
23. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–100 (1981).
24. Liu, Y. & Dilger, J. P. Opening rate of acetylcholine receptor channels. *Biophys. J.* **60**, 424–432 (1991).
25. Spitzmaul, G., Dilger, J. P. & Bouzat, C. The noncompetitive inhibitor quinuclidine modifies the desensitization kinetics of muscle acetylcholine receptors. *Mol. Pharmacol.* **60**, 235–243 (2001).
26. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
27. Sine, S. M., Wang, H.-L. & Bren, N. Lysine scanning mutagenesis delineates structural model of the nicotinic receptor ligand binding domain. *J. Biol. Chem.* **277**, 29210–29223 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health, UNS, ANPCyT and F. Antorchas, Argentina.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S. M. S. (sine@mayo.edu).

Identification of an antimalarial synthetic trioxolane drug development candidate

Jonathan L. Vennerstrom¹, Sarah Arbe-Barnes², Reto Brun³, Susan A. Charman⁴, Francis C. K. Chiu⁴, Jacques Chollet³, Yuxiang Dong¹, Arnulf Dorn⁵, Daniel Hunziker⁵, Hugues Matile⁵, Kylie McIntosh¹, Maniyan Padmanilayam¹, Josefine Santo Tomas³, Christian Scheurer³, Bernard Scoreaux³, Yuanqing Tang¹, Heinrich Urwyler⁶, Sergio Wittlin³ & William N. Charman⁴

¹College of Pharmacy, University of Nebraska Medical Center, 986025 Nebraska Medical Center, Omaha, Nebraska 68198-6025, USA

²Fulcrum Pharma Developments Ltd, Hemel Hempstead, Hertfordshire HP1 1JY, UK

³Swiss Tropical Institute, Socinstrasse 57, CH-4002 Basel, Switzerland

⁴Victorian College of Pharmacy, Monash University, 381 Royal Parade, Parkville, Victoria 3052, Australia

⁵F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

⁶Basilea Pharmaceutica Ltd, Grenzacherstrasse 487, CH-4058 Basel, Switzerland

The discovery of artemisinin more than 30 years ago provided a completely new antimalarial structural prototype; that is, a molecule with a pharmacophoric peroxide bond in a unique 1,2,4-trioxane heterocycle¹. Available evidence^{2–4} suggests that artemisinin and related peroxidic antimalarial drugs exert their parasitocidal activity subsequent to reductive activation by haem, released as a result of haemoglobin digestion by the malaria-causing parasite. This irreversible redox reaction produces carbon-centred free radicals, leading to alkylation of haem⁵ and proteins (enzymes)⁶, one of which—the sarcoplasmic-endoplasmic reticulum ATPase PfATP6 (ref. 7)—may be critical to parasite survival. Notably, there is no evidence of drug resistance to any member of the artemisinin family of drugs⁸. The chemotherapy of malaria has benefited greatly from the semi-synthetic artemisinins artemether and artesunate as they rapidly reduce parasite burden, have good therapeutic indices and provide for successful treatment outcomes⁹. However, as a drug class, the artemisinins suffer from chemical¹⁰ (semi-synthetic availability, purity and cost), biopharmaceutical¹¹ (poor bioavailability and limiting pharmacokinetics) and treatment^{8,11} (non-compliance with long treatment regimens and recrudescence) issues that limit their therapeutic potential. Here we describe how a synthetic peroxide antimalarial drug development candidate was identified in a collaborative drug discovery project.

The drug discovery process relied on prospective multidimensional lead optimization made possible by rapid and iterative integration of antimalarial, physicochemical, metabolism, pharmacokinetic and toxicity data that guided the medicinal chemistry. In comparison to the semi-synthetic artemisinins (Fig. 1a), the selected 1,2,4-trioxolane drug development candidate (trioxolane 7; also known as OZ277 or RBx-11160, Fig. 1e) exhibits structural simplicity, an economically feasible and scalable synthesis, superior antimalarial activity and an improved biopharmaceutical profile. It has progressed through formal pre-clinical regulatory studies, and will be advanced into 'first-in-man' clinical studies during 2004.

The essential characteristics for a new trioxolane antimalarial drug product were identified early in the discovery process to guide compound progression, and they ultimately led to selection of a development candidate. Key features included low product costs requiring a straightforward synthesis and simple formulation approaches, a maximum 3-day treatment regimen with once-daily administration necessitating good potency and pharmacokinetic

properties, low potential for toxicity and no evidence for resistance development. On the basis of these characteristics, a candidate selection matrix¹² based on a clearly defined target product profile was used to guide the lead optimization and compound selection processes.

Dispiro-1,2,4-trioxolanes **1–4**, four of the first five target compounds that were synthesized (Fig. 1b, c), provided a foundational structure–activity relationship, and trioxolanes **3** and **4** were identified as prototypes with activity against *Plasmodium falciparum* *in vitro*¹³ and *Plasmodium berghei* *in vivo*¹⁴ comparable to that of artemether and artesunate (Fig. 1a). That trioxolanes **1** and **2** were completely inactive (Fig. 1b) suggests that antimalarial activity decreases when the peroxide bond is too exposed, or sterically inaccessible, to iron(II) species. For trioxolanes **3** and **4**, a middle ground seems to have been met as one side of the trioxolane heterocycle is sterically hindered, but the other allows for an energetically favourable approach of iron(II) to a relatively sterically unhindered peroxide oxygen atom. Evidence to support this hypothesis was provided by the iron(II)-mediated fragmentation of trioxolane **4** by exclusive β -scission of the spiroadamantane ring to form the carbon-centred secondary free radical that was trapped by the stable nitroxide free radical 2,2,6,6-tetramethyl-1-piperidine-1-oxyl (TEMPO) to form the corresponding aminoxy acid (Fig. 1c). Because we obtained no products resulting from β -scission of the spirocyclohexanone, regioselective¹⁵ formation of the Fe(III)-complexed oxy radical resulting from attack of the Fe(II) species on the less-hindered peroxide bond oxygen atom is inferred.

It was soon evident that poor aqueous solubility and low oral bioavailability were significant issues for prototypes **3** and **4**, as the spiroadamantane trioxolane pharmacophore is inherently lipophilic. It is generally accepted that highly lipophilic compounds are more susceptible to metabolism and have higher clearance values and larger volumes of distribution^{16,17}. Indeed, trioxolane **4** and comparably lipophilic analogues displayed low aqueous solubility ($<1 \mu\text{g ml}^{-1}$) and low oral bioavailability ($<1\%$) in rats.

To identify a trioxolane development candidate, a number of structured, prospective and iterative strategies ensured efficient multidimensional lead optimization. First, we prospectively predicted relevant physicochemical properties (Log P or Log D for ionizable compounds, and polar surface area¹⁸) to reduce the risk of poor oral absorption owing to poor membrane permeability or poor aqueous solubility^{19,20}.

Second, it was necessary to define the chemistry enabling synthesis of functionalized trioxolanes with better physicochemical and biopharmaceutical properties. To avoid the synthesis of chiral analogues of trioxolane prototype **3**, we chose to use 4-substituted cyclohexanones, as only two achiral *cis* and *trans* diastereomer reaction products are possible in the cozonolysis reaction with the symmetrical *O*-methyl 2-adamantanone oxime. Two key trioxolane intermediates (Fig. 1d) were readily prepared using the Griesbaum cozonolysis reaction²¹ in ≥ 100 mmol scale with yields of 68% to 78%, and were isolated without using chromatography. X-ray crystallographic structures of trioxolane **5** and the phthalimide precursor of trioxolane **6** (data not shown) revealed that the peroxide bond is axial and the substituent is equatorial, establishing the stereochemistry as *cis*. As illustrated by trioxolanes **5–7** (Fig. 1e), a number of relatively polar target trioxolanes were synthesized using a wide range of post-ozonolysis transformations.

Third, with increasing activity of the trioxolanes, the single-dose *in vivo* antimalarial primary activity screens were refined to improve their discriminatory ability, and the criteria for progression of a trioxolane candidate from the primary screen to the more labour-intensive secondary screens became more stringent.

Fourth, we incorporated rapid *in vitro* metabolism screening of active trioxolanes using human, rat and mouse hepatic microsomes, which provided a necessary hurdle for progression of candidates to the more labour- and analytically intensive oral bioavailability and intravenous pharmacokinetic and toxicity studies. Use of a non-solubilizing, standard suspension vehicle (SSV) for assessment of oral activity ensured active candidates would have reasonable oral

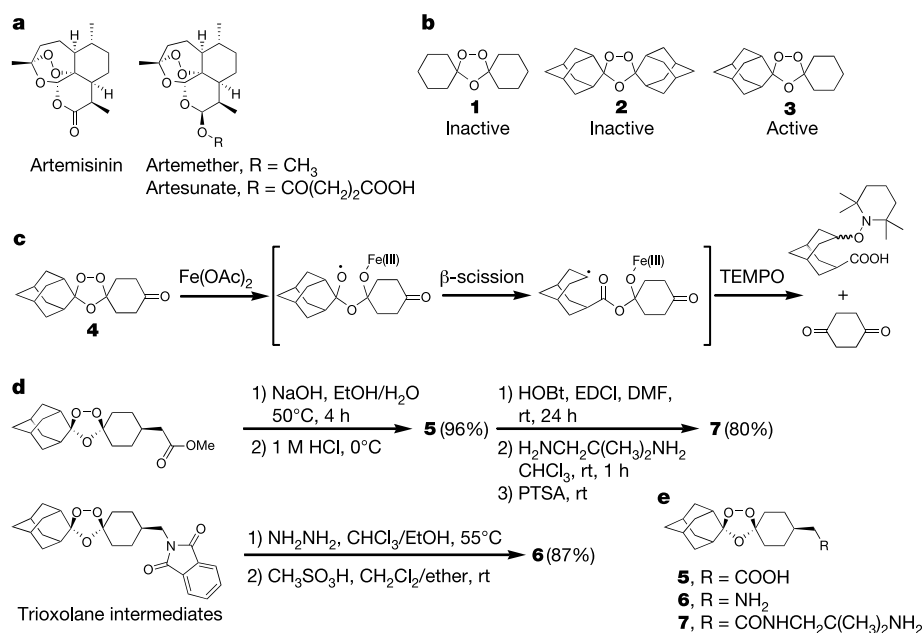


Figure 1 Trioxolane chemistry. **a**, Structures of artemisinin, artemether and artesunate. **b**, Structures of trioxolanes **1**, **2** and **3**. **c**, The reaction of trioxolane **4** with ferrous acetate and trapping of the secondary carbon radical with TEMPO. A parallel result was described previously¹⁵ in a study of the Fe(II) decomposition of trioxolanes derived from cyclopentene and indene. We note that steric hindrance affecting peroxide bond

accessibility is only one of many possibly significant factors influencing trioxolane antimalarial activity. **d**, Key trioxolane intermediates and representative post-ozonolysis transformations. rt, room temperature. **e**, Structures of trioxolanes **5–7**. Trioxolane **6** was isolated as the mesylate, trioxolane **7** (OZ277) as the tosylate.

Table 1 Half-life and bioavailability values after a single oral dose to healthy rats

Compound	Dose (mg kg ⁻¹)	Half life (h)	Bioavailability (%)
Trioxolane 5	50.0	2.0 ± 0.3 (n = 3)	74.1 ± 19.1 (n = 3)
Trioxolane 6	18.8	1.8, 1.6‡	27.5, 17.3‡
Trioxolane 7	17.4	1.4 ± 0.2 (n = 3)	35.0 ± 6.8 (n = 3)
Artemether	50.0	ND	1.4 ± 0.6 (n = 3)
Artesunate	10.0	0.47, 0.48*‡	23.3, 32.3‡

Values are mean ± s.d.

*Half life for dihydroartemisinin after dosing with artesunate.

‡Oral bioavailability based on artesunate concentrations.

‡Only two measurements available.

bioavailability, and it provided an early indication that sophisticated (and expensive) formulation approaches would not be required.

A breakthrough occurred when trioxolane **5** was found to exhibit high oral bioavailability and an acceptable half-life in rats (Table 1), demonstrating that the required pharmacokinetic characteristics could be obtained with the spiroadamantane trioxolane pharmacophore. Importantly, the pharmacokinetics of trioxolane **5** in rats represented a substantial improvement over the comparator compounds artesunate and artemether. Whereas trioxolane **5** suffers from poor intrinsic antimalarial potency (Table 2), its reasonable activity in the *P. berghei* model is probably due to its excellent pharmacokinetic properties. Trioxolane **6** provided a further landmark, combining excellent antimalarial activity with good oral bioavailability in rats (Tables 1 and 2). Trioxolane **6** was the first exception to the general trend, also exhibited by the artemisinins and other synthetic antimalarial peroxides²², where increased lipophilicity typically leads to higher antimalarial activity albeit with compromised biopharmaceutical properties.

The respective carboxylic acid and primary amine functional groups of trioxolanes **5** and **6** allowed us to use a wide range of chemical transformations to synthesize numerous derivatives^{23,24}, among which was trioxolane **7**, the selected development candidate. In the remainder of the paper, we will profile trioxolanes **6** and **7**, and disclose the particular properties of trioxolane **7** that led to its selection as the development candidate.

Against *P. falciparum* *in vitro*, trioxolanes **6** and **7** were no less potent than artemether, the most active comparator drug (Table 2). After a single 3 mg kg⁻¹ oral dose in the murine *P. berghei* model, trioxolanes **6** and **7** were clearly more active than artesunate, artemether, chloroquine and mefloquine (Table 2), a result confirmed by the effective dose experiments (Fig. 2a). In accord with our objective to identify an orally active compound that could provide a cure within 3 days of treatment, repeated dose experiments were implemented (Supplementary Table 1). At 3 × 10 mg kg⁻¹, the activities of the two trioxolanes and other comparator drugs, with the exception of artesunate, were nearly indistinguishable, but only trioxolanes **6** and **7** were curative, which

clearly differentiates the trioxolanes from the comparator drugs (Table 2).

Further secondary screening using the *P. berghei* murine model included onset and recrudescence (Fig. 2b) and prophylaxis (Supplementary Table 2) experiments. Artesunate, artemether, chloroquine and trioxolanes **6** and **7** reduced parasitaemia to below quantifiable limits (1 parasite in 10,000 erythrocytes) within 24 h, whereas the onset of action for mefloquine was considerably slower, consistent with previous reports¹⁴. For trioxolane **7** and all comparator drugs, recrudescence occurred between day 6 and 14 after infection, whereas for trioxolane **6**, there was no recrudescence and all mice were cured. In prophylaxis experiments, mefloquine and trioxolane **6** had similar profiles with complete protection when administered 24 h before infection and high protection when given at 72 h before infection. Chloroquine protected mice when administered 24 h before infection, whereas trioxolane **7**, artesunate and artemether had little or no prophylactic activity.

Values of half-maximal inhibitory concentration (IC₅₀) determined against ten *P. falciparum* strains from different geographical regions differed four- to sixfold for trioxolanes **6** and **7**, and sevenfold for artesunate, suggesting that cross-resistance is an unlikely problem for these trioxolanes (Supplementary Table 3). Profiles for trioxolanes **6** and **7** in combination with chloroquine, pyrimethamine and mefloquine against the NF54 and K1 strains of *P. falciparum* (data not shown) were similar to those described earlier for parallel combinations of artesunate and artemether with these drugs²⁵.

In vitro metabolism studies assessed the susceptibility of compounds to cytochrome P450 (CYP450)-mediated metabolism and the possibility of drug–drug interactions. After incubation with human hepatic microsomes and cryopreserved human hepatocytes, there was minimal loss of trioxolanes **6** and **7**, indicative of a low susceptibility for metabolism by CYP450 enzymes (data not shown). Additional *in vitro* studies conducted to determine the potential for CYP450 inhibition indicated that trioxolane **6** moderately inhibited CYP3A4 and CYP2C9 and strongly inhibited CYP2C19, whereas trioxolane **7** minimally inhibited CYP3A4, CYP1A2, CYP2C9 and CYP2D6, and only moderately inhibited CYP2C19.

The significant advantages of lead candidates **6** and **7** in terms of half-life and oral bioavailability can be seen in Table 1. After intravenous administration, plasma concentrations of both trioxolanes **6** and **7** were measurable for up to 6 h, and terminal elimination half-lives were significantly longer than for dihydroartemisinin, the major (active) metabolite of artesunate²⁶. After oral administration as an SSV formulation to rats, the absolute bioavailability for trioxolanes **6** and **7** greatly exceeded that for artemether and were comparable to artesunate. At this dose, plasma concentrations for both trioxolane **6** and **7** remained above 10 ng ml⁻¹ for at least 8 h (Supplementary Fig. 1), whereas concentrations of

Table 2 *In vitro* activity against *P. falciparum* and *in vivo* activity in *P. berghei*-infected mice

Compound	IC ₅₀ (ng ml ⁻¹) *		1 × 3 mg kg ⁻¹ (oral)		3 × 10 mg kg ⁻¹ (oral)		
	Strain K1†	Strain NF54‡	Activity (%)	Survival (days)	Activity (%)	Survival (days)	Cure (%)§
Control	—	—	0	5.2	0	5.2	0
Trioxolane 5	34 ± 6	45 ± 6	50	9.0	NT	—	—
Trioxolane 6	0.39 ± 0.06	0.42 ± 0.06	99	10.0	>99.99	30.0	100
Trioxolane 7	1.0 ± 0.1	0.91 ± 0.12	98	9.0	>99.99	26.2	67
Artesunate	1.3 ± 0.2	1.6 ± 0.1	33	6.6	97	11.0	0
Artemether	0.74 ± 0.11	1.2 ± 0.1	56	8.0	>99.99	22.3	0
Chloroquine	62 ± 4	5.1 ± 0.8	85	7.9	99.99	18.2	0
Mefloquine	3.0 ± 0.1	5.8 ± 0.2	18	7.0	99.92	24.3	0

NT, not tested.

*Mean ± s.e.m. (n ≥ 10).

†Chloroquine-resistant (Thailand).

‡Chloroquine-sensitive (airport, unknown origin).

§No detectable parasites at 30 days after infection.

artemether and dihydroartemisinin were below quantifiable limits in less than 3 h (even when administered at significantly higher doses). Additional *in vivo* studies conducted in beagle dogs (data not shown) were consistent with the rat data for trioxolanes 6 and 7.

The whole-blood concentrations of trioxolanes 6 and 7 in rats exceeded plasma concentrations by approximately 2.5:1 throughout the post-dosing period after oral administration. *In vitro* studies confirmed a similar whole blood to plasma ratio in freshly collected healthy mouse blood; however, there was a substantially increased ratio ($>100:1$) in blood taken from mice infected with *P. berghei* at a level of approximately 30% parasitaemia. Partitioning in parasite-infected mouse blood was concentration dependent, as it was considerably higher at lower total blood concentrations, thereby indicating a possible saturable process. A similar trend for preferential and concentration-dependent uptake into infected erythrocytes was recently reported for artemisinin²⁷. Therefore, concentrations of trioxolanes 6 and 7 within infected erythrocytes are likely to be substantially higher than, and not directly proportional to, concentrations measured in plasma or whole blood. This has important implications regarding the relevance of measured blood or plasma drug concentrations, as it is likely that

they will significantly underestimate concentrations at the site of action.

Preliminary studies conducted to assess tissue distribution after oral administration to rats indicated that at 2 h after dosing, trioxolane 6 was extensively distributed throughout the major organs with significant concentrations detected in liver, kidney, brain, lung and heart tissue (Fig. 3a). At 2 h after dosing, tissue concentrations of trioxolane 7 were two- to fivefold lower than those of trioxolane 6, and there was no measurable concentration of trioxolane 7 in brain tissue. At 18 h after dosing, measurable concentrations of trioxolane 6 were present in brain and lung tissue, but trioxolane 7 was quantifiable only in lung tissue (Fig. 3b). The notable absence of quantifiable concentrations of trioxolane 7 in brain tissue over the 18-h sampling period was considered a possible advantage given the concern for potential neurotoxicity with the semi-synthetic artemisinins²⁸.

No relevant mutagenic potential was identified for trioxolanes 5–7 using a miniaturized version of the standard Ames test for bacterial gene mutations, or a test for chromosomal aberrations in mammalian cells using a murine lymphoma cell line. Multiple dose (5 day) oral toxicity testing in male Wistar rats indicated that the toxicological profiles of trioxolanes 5–7 were comparable to that of artesunate with only quantitative differences, and included minor clinical findings, gastric irritation, hepatocellular hypertrophy, renal tubular changes, and atrophy of lymphatic tissue at high doses (for example, 300 mg kg^{-1}). Trioxolane 7 ranked intermediate between trioxolanes 5 and 6 with an overall acceptable toxicity profile. Findings were essentially reversible after a 1-week recovery period and there was no indication of neurotoxicity with any of the trioxolanes tested. Analysis of plasma samples indicated that

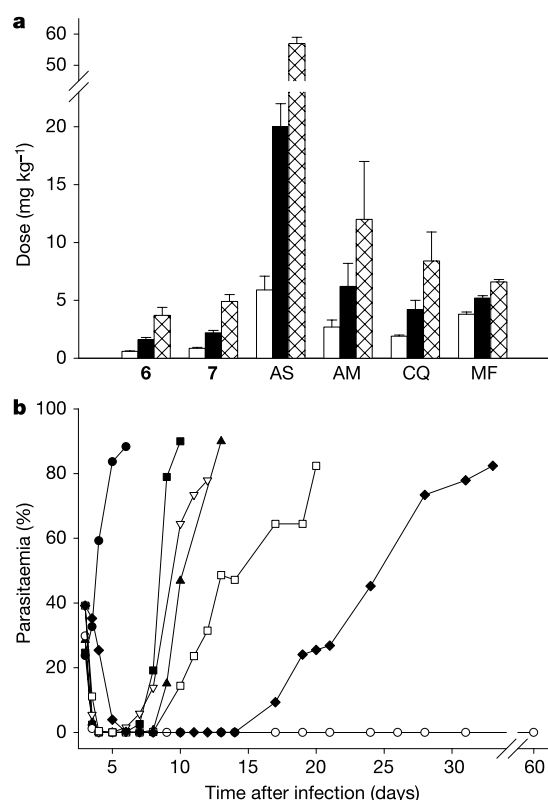


Figure 2 Effective dose and onset and recrudescence data. **a**, Effective doses resulting in 50% (open bars), 90% (filled bars) and 99% (hatched bars) decrease in parasitaemia relative to control values. Values were determined after treatment of $n = 5$ mice with a single dose on day 1 after infection with blood smears taken on day 3 (72 h after infection) for determination of effective dose values. Data are shown as mean $\pm$ s.e.m. with data from at least two independent experiments. AM, artemether; AS, artesunate; CQ, chloroquine; MF, mefloquine. **b**, Onset of action and recrudescence after a single oral dose of 100 mg kg^{-1} to a group of five mice on day 3 after infection. Parasitaemia reduction was monitored initially at 12 h after treatment, and the time of recrudescence was assessed by daily blood smears followed by intermittent assessment for up to 60 days. Control (filled circles), artesunate (filled squares), artemether (open triangles), chloroquine (open squares), mefloquine (filled diamonds), trioxolane 6 (open circles) and trioxolane 7 (filled triangles) are shown.

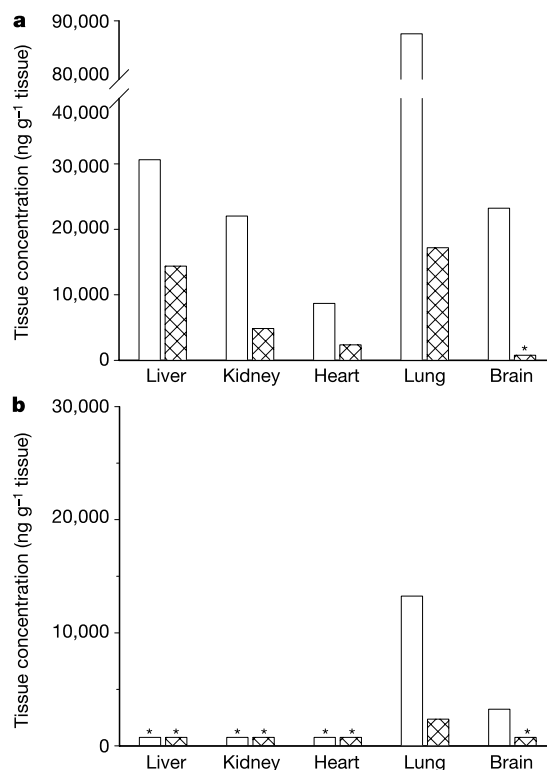


Figure 3 Tissue concentrations of trioxolanes 6 (open bars) and 7 (hatched bars) after oral administration of approximately 35 mg kg^{-1} to rats. Tissues were collected at 2 h (**a**) and 18 h (**b**) after dosing, immediately placed on ice and a 1–2 g portion homogenized in water to give a 0.2 g ml^{-1} suspension. An aliquot was extracted by the addition of acetonitrile (2:1 v/v) and analysed by LC/MS. Asterisks indicate that the concentration was below the limit of quantification of the assay (that is, $<750 \text{ ng g}^{-1}$ tissue).

trioxolanes **5**, **6** and **7** were present at high concentrations after oral dosing ($>500\text{--}1,000\text{ ng ml}^{-1}$ after 5 days dosing). The selection of trioxolane **7** as the development candidate over trioxolane **6**, in spite of the compelling activity of trioxolane **6**, was primarily based on its improved toxicological profile and evidence for reduced concentrations in brain tissue after oral dosing. □

Methods

Synthesis of trioxolanes 1–7

Trioxolanes 1–7 were synthesized using the Griesbaum cozonolysis²¹ and post-ozonolysis reactions as previously described^{23,24,29}. Differential scanning calorimetry experiments revealed that the trioxolanes had good thermal stability with decomposition enthalpies ranging from 400 to 700 J g^{−1} compared with values over 1,000 J g^{−1} for artemether and standard endoperoxides.

In vitro metabolism

Compounds were incubated with human liver microsomes (BD Gentest, Discovery Labware Inc.) at a substrate concentration of 1–10 μM and a microsomal protein concentration of 0.4 mg ml^{−1}. Incubations were conducted at 37 °C for 60 min in 0.1 M phosphate buffer (pH 7.4) containing 1 mg ml^{−1} NADP⁺, 1 mg ml^{−1} glucose-6-phosphate, 1 U ml^{−1} glucose-6-phosphate dehydrogenase and 0.67 mg ml^{−1} MgCl₂. Controls were included without addition of cofactor. At selected time points, samples and controls were quenched by the addition of acetonitrile (2:1 v/v), filtered and assayed by LC/MS. The enzyme inhibition assay was based on a previously published method³⁰ with minor modifications.

In vivo pharmacokinetic studies in rats

In vivo studies were approved by the institutional animal experimentation ethics committee. Intravenous formulations included 0.1 M Captisol (CyDex Inc) in water (for artemether), normal saline (for artesunate), isotonic pH 3.0 citrate buffer (for trioxolane **6**), and 7.5% DMSO in pH 3.0 citrate buffer (for trioxolane **7**). Intravenous doses were administered by infusion via a cannula surgically implanted in the jugular vein on the day before dosing. For oral dosing, all compounds were formulated in SSV (0.5% sodium carboxymethylcellulose, 0.5% benzyl alcohol, 0.4% Tween 80 in 0.9% NaCl) and administered to fasted rats via oral gavage. Blood samples (0.25 ml) were collected from a cannula surgically implanted in the carotid artery on the day before dosing and were immediately centrifuged and the plasma separated and frozen at −20 °C. Before analysis by LC/MS, plasma proteins were precipitated by the addition of acetonitrile (2:1 v/v). Quantification was conducted by comparison to a calibration curve prepared using blank plasma and processed in the same way as the samples. Bioavailability was determined using the area under the plasma concentration versus time (AUC) profiles extrapolated to infinity. Truncated AUC values were used in cases where the terminal rate constant could not be determined.

LC/MS analysis

Samples were assayed by LC/MS using either a Micromass Quattro Ultima Pt triple quadrupole instrument coupled with a Waters 2795 HPLC (artesunate/dihydroartemisinin, trioxolane **7**) or a Micromass ZQ single quadrupole instrument coupled to a Waters 2690 HPLC (artemether and trioxolanes **5** and **6**). Columns included a Phenomenex Luna C8(2) (5 μm, 50 × 2 mm internal diameter) for trioxolane **7** and an SGE C18(2) (5 μm, 50 × 2 mm internal diameter) for artemether, artesunate/dihydroartemisinin and trioxolanes **5** and **6**, each being maintained at 50 °C. The mobile phase consisted of formic acid, acetonitrile and water (trioxolanes **6**, **7**), ammonium formate, acetonitrile and water (artesunate/dihydroartemisinin), or ammonium acetate, acetonitrile and water (artemether), and all compounds were eluted under gradient conditions. Mass spectrometry was conducted under ESI conditions with detection by selected ion monitoring (artemether and trioxolanes **5** and **6**) or multiple reactions monitoring (artesunate/dihydroartemisinin and trioxolane **7**). Assays were suitably validated for linearity, accuracy, reproducibility and compound recovery. Limits of quantification were 50 ng ml^{−1} (artemether), 20 ng ml^{−1} (artesunate/dihydroartemisinin), 10 ng ml^{−1} (trioxolane **5**) and 1 ng ml^{−1} (trioxolanes **6** and **7**).

In vitro antimalarial activity

In vitro antimalarial activity was measured using the [³H]-hypoxanthine incorporation assay¹³ with various strains of *P. falciparum* obtained from Roche. Results were expressed as the concentration resulting in 50% inhibition (IC₅₀).

In vivo antimalarial efficacy studies

All efficacy studies were approved by the institutional animal experimentation ethics committee. *In vivo* antimalarial activity was usually assessed for groups of three–five female NMRI mice (20–22 g) intravenously infected on day 0 with *P. berghei* strain ANKA (2 × 10⁷ parasitized erythrocytes per ml) (Roche). In control mice, parasitaemia typically rose to ~40% by day 3 after infection, and control mice died between day 5 and day 7 after infection. Compounds formulated in SSV were administered orally in a volume of 10 ml kg^{−1} either as a single dose (24 h after infection) or as three consecutive daily doses (24, 48 and 72 h after infection). With the single-dose regimen, blood smears were collected on day 3 (72 h after infection) and stained with Giemsa. The blood smears for the triple-dose regimens were collected and stained on day 4 after infection. The degree of infection (parasitaemia expressed as per cent infected erythrocytes) was determined microscopically, with a detection limit of 1 parasite in 10,000 erythrocytes (that is, 0.01%).

Activity was calculated as the difference between the mean per cent parasitaemia for the control and treated groups expressed as a per cent relative to the control group. The survival time in days was also recorded up to 30 days after infection. A compound was considered curative if the animal survived to day 30 after infection with no detectable parasites.

Received 31 March; accepted 18 June 2004; doi:10.1038/nature02779.

- Klayman, D. L. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228**, 1049–1055 (1985).
- Jefford, C. W. Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. *Curr. Med. Chem.* **8**, 1803–1826 (2001).
- Cumming, J. N., Polypradith, P. & Posner, G. H. Antimalarial activity of artemisinin (qinghaosu) and related trioxanes: mechanism(s) of action. *Adv. Pharmacol.* **37**, 254–297 (1997).
- Wu, Y. How might qinghaosu (artemisinin) and related compounds kill the intraerythrocytic malaria parasite? A chemist's view. *Acc. Chem. Res.* **35**, 255–259 (2002).
- Robert, A., Coppel, Y. & Meunier, B. Alkylation of heme by the antimalarial drug artemisinin. *Chem. Commun.*, 414–415 (2002).
- Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity. *Int. J. Parasitol.* **32**, 1655–1660 (2002).
- Eckstein-Ludwig, U. *et al.* Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature* **424**, 957–961 (2003).
- White, N. J. Antimalarial drug resistance and combination chemotherapy. *Phil. Trans. R. Soc. Lond. B* **354**, 739–749 (1999).
- White, N. J. Assessment of the pharmacodynamic properties of antimalarial drugs *in vivo*. *Antimicrob. Agents Chemother.* **41**, 1413–1422 (1997).
- Avery, M. A., Chong, W. K. M. & Jennings-White, C. Stereoselective total synthesis of (+)-artemisinin, the antimalarial constituent of *Artemisia annua* L. *J. Am. Chem. Soc.* **114**, 974–979 (1992).
- Ridley, R. G. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* **415**, 686–693 (2002).
- Nwaka, S. & Ridley, R. G. Virtual drug discovery and development of neglected diseases through public-private partnerships. *Nature Rev. Drug Discov.* **2**, 919–928 (2003).
- Desjardins, R. E., Canfield, C. J., Haynes, J. D. & Chulay, J. D. Quantitative assessment of antimalarial activity *in vitro* by a semiautomated microdilution technique. *Antimicrob. Agents Chemother.* **16**, 710–718 (1979).
- Ridley, R. G. *et al.* Antimalarial activity of the bisquinoline *trans*-N¹, N²-bis-(7-chloroquinolin-4-yl)cyclohexane-1,2-diamine: Comparison of two stereoisomers and detailed evaluation of the S,S enantiomer, Ro 47-7737. *Antimicrob. Agents Chemother.* **41**, 677–686 (1997).
- Abe, M., Inakazu, T., Munakata, J. & Nojima, M. ¹⁸O-Tracer studies of Fe(II)-induced decomposition of 1,2,4-trioxolanes (ozonides) derived from cyclopentenones and indenones. Inner-sphere electron transfer reduction of the peroxide linkage. *J. Am. Chem. Soc.* **121**, 6556–6562 (1999).
- van de Waterbeemd, H., Smith, D. A., Beaumont, K. & Walker, D. K. Property-based design: Optimisation of drug absorption and pharmacokinetics. *J. Med. Chem.* **44**, 1313–1333 (2001).
- Smith, D., Jones, B. C. & Walker, D. K. Design of drugs involving concepts and theories of drug metabolism and pharmacokinetics. *Med. Res. Rev.* **16**, 243–266 (1996).
- Palm, K., Stenberg, P., Luthman, K. & Artursson, P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **14**, 568–571 (1997).
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–26 (1997).
- Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
- Griesbaum, K., Liu, X., Kassiaris, A. & Scherer, M. Ozonolyses of O-alkylated ketoximes in the presence of carbonyl groups: a facile access to ozonides. *Liebigs Ann./Recueil*, 1381–1390 (1997).
- Tang, Y., Dong, Y. & Vennerstrom, J. L. Synthetic peroxides as antimalarials. *Med. Res. Rev.* **24**, 425–448 (2004).
- Vennerstrom, J. L., Dong, Y., Chollet, J. & Matile, H. Spiro and dispiro 1,2,4-trioxolane antimalarials. US patent 6, 486,199 (2002).
- Vennerstrom, J. L. *et al.* Spiro and dispiro 1,2,4-trioxolane antimalarials. US continuation-in-part based on PCT/US02/19767 (filed 21 June 2002).
- Chawira, A. N. & Warhurst, D. C. The effect of artemisinin combined with standard antimalarials against chloroquine-sensitive and chloroquine-resistant strains of *Plasmodium falciparum* *in vitro*. *J. Trop. Med. Hyg.* **90**, 1–8 (1987).
- Li, A.-G. *et al.* The pharmacokinetics and bioavailability of dihydroartemisinin, arteether, artemether, artesunate and artemisinic acid in rats. *J. Pharm. Pharmacol.* **50**, 173–182 (1998).
- Vyas, N., Avery, B. A., Avery, M. A. & Wyandt, C. M. Carrier-mediated partitioning of artemisinin into *Plasmodium falciparum*-infected erythrocytes. *Antimicrob. Agents Chemother.* **46**, 105–109 (2002).
- Gordi, T. & Lepist, E.-I. Artemisinin derivatives: Toxic for laboratory animals, safe for humans? *Toxicol. Lett.* **147**, 99–107 (2004).
- Keul, H. Über konstitution und entstehung der ozonide von bis-adamantyliden und von bis-bicyclo[3.3.1]non-9-yliden. *Chem. Ber.* **108**, 1207–1217 (1975).
- Crespi, C. L., Miller, V. P. & Penman, B. W. Microtiter plate assays for inhibition of human, drug-metabolizing cytochromes P450. *Anal. Biochem.* **248**, 188–190 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. G. Ridley and M. Tanner for inspiration; K. Griesbaum and A. Hudson for advice; C. Craft, S. Nwaka, S. Campbell, P. Hadvary and R. Imhof for their support; and J. M. Karle for performing X-ray crystallographic experiments. This work was supported by the World Health Organization and Medicines for Malaria Venture.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L.V. (jvenners@unmc.edu).

Crystal structure of a complex between anthrax toxin and its host cell receptor

Eugenio Santelli¹, Laurie A. Bankston¹, Stephen H. Leppla² & Robert C. Liddington¹

¹Program on Cell Adhesion, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

²Microbial Pathogenesis Section, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20892, USA

Anthrax toxin consists of the proteins protective antigen (PA), lethal factor (LF) and oedema factor (EF)¹. The first step of toxin entry into host cells is the recognition by PA of a receptor on the surface of the target cell. Subsequent cleavage of receptor-bound PA enables EF and LF to bind and form a heptameric PA₆₃ prepore, which triggers endocytosis. Upon acidification of the endosome, PA₆₃ forms a pore that inserts into the membrane and translocates EF and LF into the cytosol². Two closely related host cell receptors, TEM8 and CMG2, have been identified. Both bind to PA with high affinity and are capable of mediating toxicity^{3,4}. Here, we report the crystal structure of the PA–CMG2 complex at 2.5 Å resolution. The structure reveals an extensive receptor–pathogen interaction surface mimicking the non-pathogenic recognition of the extracellular matrix by integrins⁵. The binding surface is closely conserved in the two receptors and

across species, but is quite different in the integrin domains, explaining the specificity of the interaction. CMG2 engages two domains of PA, and modelling of the receptor-bound PA₆₃ heptamer^{6–8} suggests that the receptor acts as a pH-sensitive brace to ensure accurate and timely membrane insertion. The structure provides new leads for the discovery of anthrax anti-toxins, and should aid the design of cancer therapeutics⁹.

Both TEM8 and CMG2 contain a domain that is homologous to the I domains of integrins, which comprise a Rossman-like α/β -fold with a metal-ion-dependent adhesion site (MIDAS) motif on their upper surface¹⁰. Crystal structures of the CMG2 I domain and full-length PA proteins have previously been determined^{6,11}. The PA monomer is a long slender molecule comprising four distinct domains. In the PA–CMG2 I domain complex, two of these four domains (II and IV) pack together at the base of PA and engage the upper surface of the CMG2 I domain surrounding the MIDAS motif (Fig. 1), burying a large protein surface (1,900 Å²), consistent with the very high affinity (sub-nanomolar dissociation constant) of this interaction¹². The I domain adopts the ‘open’ conformation, typical of integrin–ligand complexes^{5,13}. PA mimics the ligand recognition mechanism of the integrins⁵ by contributing an aspartic acid side chain that completes the coordination sphere of the MIDAS magnesium ion, as predicted by mutagenesis^{14,15} (Fig. 2a, b). This single interaction contributes substantially to binding, as mutation of the aspartic acid to asparagine completely eliminates toxicity, as does mutation of a metal-coordinating residue on the receptor.

However, the MIDAS bond does not fully explain the specificity of the interaction, as it does not distinguish between CMG2 and integrins. Further specificity arises from two additional interactions. First, PA domain IV docks onto the surface of CMG2 adjacent to the MIDAS motif. Domain IV comprises a β -sandwich with an immunoglobulin-like fold, but the mode of binding is quite different from that of antibody–antigen recognition. One of the receptor loops ($\alpha 2$ – $\alpha 3$) emanating from the MIDAS motif forms a hydrophobic ridge that inserts into a groove formed by one edge of the β -sandwich where its hydrophobic core is exposed. Flanking this ridge-in-groove arrangement are two further loops from CMG2, which make a number of specific polar interactions and salt bridges (Figs 3 and 4a). Together with the MIDAS contact, CMG2 and PA domain IV bury 1,300 Å² of surface area, a value very similar to two integrin–ligand interactions that have affinities in the sub-micromolar range^{5,13}. CMG2 and TEM8 share 60% identity in their I

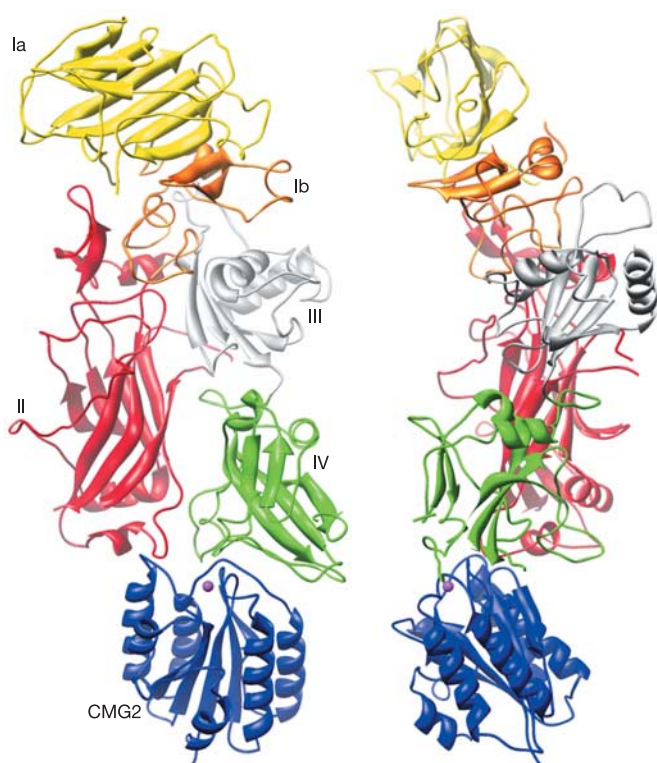


Figure 1 Structure of the PA–CMG2 complex. Two orthogonal views are shown in ribbon representation. PA is coloured by domain (I–IV). CMG2 is blue; the metal ion is shown as a magenta ball. PA domain I is cleaved after receptor binding, leading to the loss of domain Ia (yellow) and the formation of PA₆₃. All molecular graphics images were generated using the UCSF Chimera package²⁹ (<http://www.cgl.ucsf.edu/chimera>).

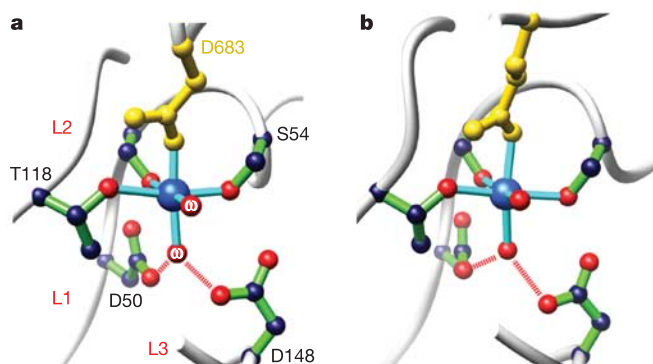


Figure 2 The MIDAS motifs of the PA–CMG2 complex (a) and the collagen–integrin $\alpha 2 \beta 1$ complex⁵ (b). Coordinating side chains and two water molecules (ω) are shown in ball-and-stick representation. The metal is shown in blue. D683 from PA, and a collagen glutamic acid, are in gold. Bond distances to the metal are 2.1 ± 0.2 Å in both cases. The three MIDAS loops (L1–L3) are labelled in a.

domains, and homology modelling based on the CMG2 structure shows that this ridge is well conserved in TEM8 and their murine counterparts, implying that they will bind PA in a similar fashion; however, the structure and sequence of the ridge are very different in integrins, explaining their weak binding (Fig. 4b).

The interaction between PA domain II and CMG2 was not anticipated. A β -hairpin from a well-ordered loop ($\beta 3$ – $\beta 4$) at the bottom of domain II inserts into a pocket on the receptor, burying 600 Å² of protein surface (Fig. 4b, c). This additional contact may explain the very high affinity of the PA–CMG2 interaction. The pocket is adjacent to the MIDAS motif and is formed by two exposed tyrosine residues (Y119 and Y158) and the $\beta 4$ – $\alpha 4$ loop, which line the sides of the pocket, and by a histidine (H121) at its base. The pocket is conserved in TEM8, but does not exist in the I domains of integrins, thus providing further specificity (Fig. 4b, c). The importance of this loop was shown by systematic mutation of the PA molecule, which revealed three mutations in this loop that reduced toxicity by >100-fold, including G342 at the tip of the β -hairpin that inserts into the pocket¹⁶.

Biophysical studies of channel conductance by PA₆₃ pores indicate that the entire region encompassed by residues 275–352 (strands $\beta 2$ and $\beta 3$ and flanking loops; see Fig. 3) in domain II rearranges to form a long β -hairpin that lines the channel lumen^{7,8}. This requires that the $\beta 2$ and $\beta 3$ strands and the $\beta 3$ – $\beta 4$ loop peel away from the side of domain II. For this to happen, domain IV, which packs against them in the pre-pore, must separate at least transiently from domain II. Thus, by binding to both domains II and IV, CMG2 may restrain the conformational changes that lead to membrane insertion. Indeed, whereas PA₆₃ heptamers insert into

artificial planar bilayers (in the absence of receptor) when the pH is reduced to 6.5, the pH requirement for receptor-mediated insertion on cells is more stringent, requiring a pH of 5.5 (ref. 17). Thus, we propose that the binding of CMG2 to the $\beta 3$ – $\beta 4$ loop stabilizes the pre-pore conformation at neutral pH; that is, the receptor may act as a brace to prevent premature membrane insertion on the cell surface before endocytosis. The pH profile of membrane insertion is consistent with the titration of histidine residues, and seven of the nine histidines within PA₆₃ cluster at the domain II–IV interface (Fig. 3). In addition, the histidine at the base of the CMG2 pocket

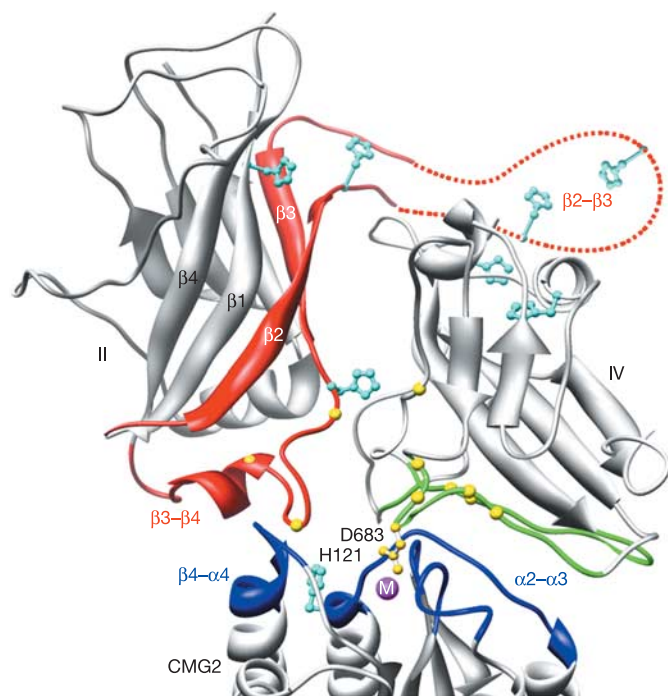


Figure 3 Intermolecular contacts between PA domains II and IV and CMG2. Contacting regions are coloured blue and green for CMG2 and PA domain IV, respectively. The $\beta 2$ – $\beta 3$ loop and flanking regions of PA domain II, which are implicated in pore formation, are highlighted in red. The $\beta 2$ – $\beta 3$ loop is disordered in monomeric PA and is shown schematically as a dashed line. The histidine residues within PA domains II and IV and within the CMG2 I domain are shown coloured cyan and are in ball-and-stick representation. Mutation sites that reduce binding by >100-fold (D683, S337, G342, W346, I656, N657, I665, Y681, N682, P686, L687) are highlighted in gold.

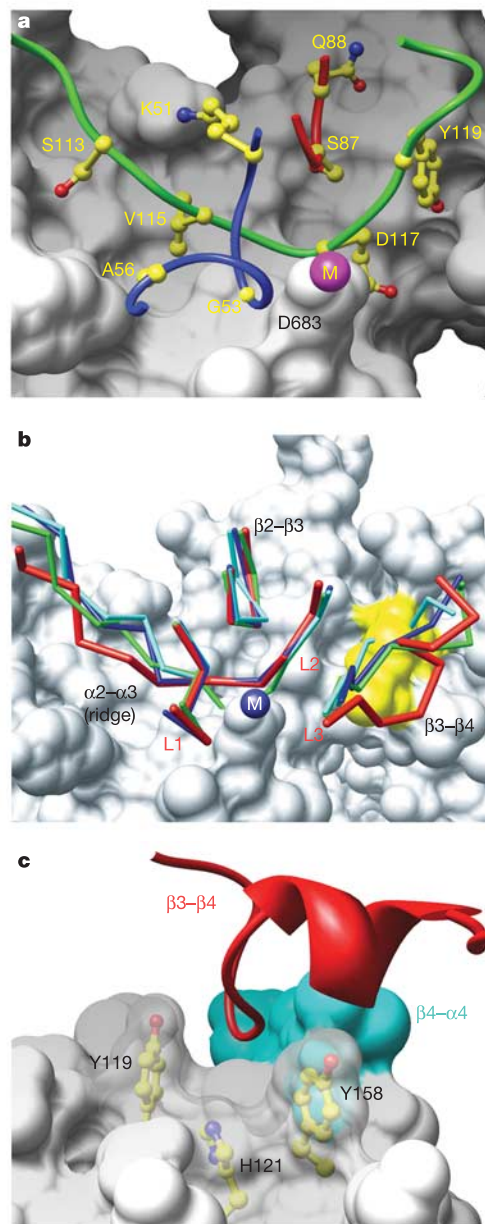


Figure 4 Key elements of the PA–CMG2 interaction **a**, Solvent-accessible surface of the PA domain IV groove, with key side chains from three CMG2 loops ($\beta 1$ – $\alpha 1$, blue; $\beta 2$ – $\beta 3$, red; $\alpha 2$ – $\alpha 3$, green) shown in ball-and-stick representation. The $\alpha 2$ – $\alpha 3$ loop forms the ridge. The MIDAS metal is labelled (M). **b**, Comparison with integrin I domains in the 'open' conformation (CMG2, red; αM , cyan; $\alpha 2$, green; αL , blue) overlaid on the MIDAS motif. **c**, Surface of the CMG2 pocket into which the PA $\beta 3$ – $\beta 4$ loop (red ribbon) inserts, formed by three CMG2 side chains (shown in ball-and-stick representation) and the $\beta 4$ – $\alpha 4$ loop (cyan).

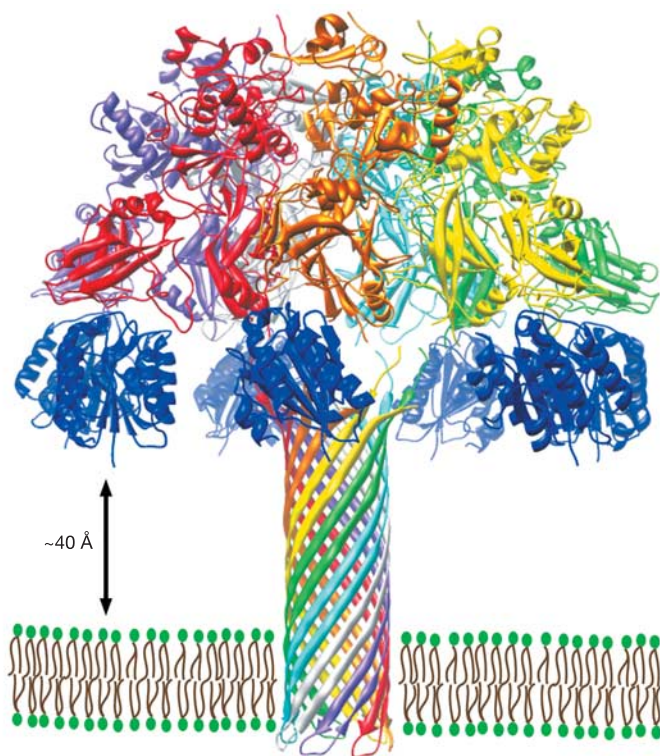


Figure 5 Hypothetical model of the receptor-bound, membrane-inserted PA pore. The model is based on the pre-pore PA₆₃ crystal structure⁶, channel conductance studies⁸, and the crystal structure of α -haemolysin¹⁹. The barrel is formed by rearrangement in each monomer of the segment shown in red in Fig. 3. Each PA₆₃ monomer is shown in a different colour. Residues 303–324 form the membrane-spanning region of the barrel. Seven copies of the CMG2 I domain bound to the heptamer are in blue. The ~40 Å gap between the CMG2 I domain and the membrane may be occupied by a ~100-residue domain of CMG2, C-terminal to the I domain, which precedes its membrane-spanning sequence.

(conserved in TEM8) has no H-bonding partners, and is close to an arginine side chain from the β 3– β 4 loop of PA. Histidine protonation provides a plausible trigger for the release of domain II from CMG2 in the acidified endosome. Indeed, we have shown that the structure of the β 3– β 4 loop is pH-sensitive, as it becomes disordered when crystals of PA grown at pH 7.5 (in the absence of receptor) are reduced to pH 6.0 (ref. 18).

It is straightforward to model the 7:7 heptameric PA₆₃–CMG2 complex, as the crystal structure of the pre-pore is known⁶ (Fig. 5). Seven CMG2 I domains lie at the base of the heptameric 'cap', increasing its height by 35 Å. The I domains are well separated, consistent with a 7:7 binding stoichiometry¹², and their amino- and carboxy termini point downwards, towards the membrane. In the transition from pre-pore to pore, the seven hairpin loops, one from each PA monomer^{6,8}, are predicted to create a 14-stranded, membrane-spanning β -barrel. Assuming an α -haemolysin-like structure¹⁹, the barrel extends ~75 Å below the I domains, with the bottom 30 Å spanning the membrane. This leaves ~40 Å between the bottom of the I domains and the membrane surface, which may be occupied by the second domain of CMG2, which comprises ~100 residues between the I domain and its C-terminal transmembrane sequence. Thus, the receptor may support the heptamer at the correct height above the membrane for accurate membrane insertion, which is stoichiometric on cells but less efficient in the absence of receptor¹⁷.

Soluble versions of the CMG2 and TEM8 I domains protect

Table 1 Data collection and refinement statistics

Parameter	Value
Space group	$P2_12_12_1$
Unit cell (Å)	$a = 88.2, b = 94.1, c = 135.6$
Resolution (Å)	30–2.5
Wavelength (Å)	0.892
R_{merge} (%)	17.6 (99.1)
I/σ	11.5 (2.4)
σ -cutoff	None
Average redundancy	5.3 (5.2)
Completeness (%)	99.9
Mosaicity	0.4
R_{work} (last shell)	20.7 (27.5)
R_{free} (last shell)	26.6 (37.2)
σ -cutoff	None
B factors (Å ²)*	32.9, 21.4, 23.3
r.m.s.d. bond lengths (Å)	0.17
r.m.s.d. bond angles (°)	1.65
Ramachandran plot (residues, %)	
Most favoured	655 86.3%
Additionally allowed	101 13.3%
Generously allowed	3 0.4%
Disallowed	0 0%

Values in parentheses refer to the highest resolution shell (2.59–2.50 Å).

*The three values are for Wilson, main chain and side chain, respectively.

against anthrax (*Bacillus anthracis*) toxin by acting as decoys^{3,15}, and our structure will allow for the design of new therapeutic agents that disrupt the PA–receptor interaction. TEM8 is strongly upregulated on the surface of endothelial cells that line the blood vessels of tumours^{20,21}, allowing for the development of anthrax toxin as an anti-tumour agent²²; however, toxicity may arise as CMG2 is expressed in most tissues. Although we expect the interactions of TEM8 and CMG2 with PA to be very similar, there are significant differences that may be exploited in the design of PA molecules that would bind better to TEM8 than to CMG2, thus minimizing the side effects from toxin binding to normal tissues. For example, V115 of CMG2, which lies at the heart of the interface with PA domain IV, is a glycine in TEM8, whereas the rim of the pocket that accepts the PA domain II loop has the sequence DGL in CMG2 but is replaced by the sequence HED in TEM8. □

Methods

Protein expression and purification

Full-length PA (residues 1–735) was prepared as previously described¹⁴. The I domain of human CMG2 was cloned as an N-terminal His-tag fusion in pET15b (Novagen) and expressed in *Escherichia coli* strain BL21(DE3). After induction of cell cultures with 0.5 mM IPTG for 2 h at 37 °C, CMG2 was purified from the soluble fraction of the cell lysate by nickel affinity chromatography (HiTrap chelating HP, Pharmacia), followed by removal of the tag with thrombin (Sigma), ion exchange (HiTrap monoQ, Pharmacia) and gel filtration (Superdex S75, Pharmacia), affinity removal of thrombin (HiTrap benzamidine FF, Pharmacia) and incubation in a buffer containing 100 mM EDTA to strip-bound metal. The final product was dialysed and concentrated to 15–20 mg ml^{−1} and flash-frozen in 150 mM NaCl, 20 mM TrisCl pH 7.5, and comprises residues 40–218 of CMG2³⁸⁶ (GenBank accession number AAK77222) plus an N-terminal extension of sequence GSHMLEDPRG as a result of the cloning strategy. The molecular mass was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. To prepare the PA–CMG2 complex, PA was mixed at a final concentration of 4 mg ml^{−1} with a threefold molar excess of CMG2 and a twofold excess of MnCl₂, incubated for 20 min at room temperature and purified by gel filtration (Superdex S200, Pharmacia). The complex was extensively dialysed and exchanged, and concentrated to 6 mg ml^{−1} in 20 mM TrisCl pH 7.5, 10 μ M MnCl₂ for crystallization trials.

Crystallization and structure solution

Needle-like crystals grew to a size of 10 × 10 × 500 μ m in 5–10 days at room temperature in a sitting-drop vapour diffusion set-up using a reservoir buffer containing 50–100 mM CHES pH 9.0–9.2, 25% PEG400. Crystals were flash-frozen at 4 °C in liquid nitrogen using the crystallization buffer with 40% PEG400 as a cryo-protectant before diffraction analysis. The crystals belong to space group $P2_12_12_1$ with unit cell parameters $a = 88.2$ Å, $b = 94.2$ Å, $c = 135.6$ Å. There is one PA–CMG2 complex in the asymmetric unit. A complete native data set to 2.5 Å was collected at beamline 9-1 at SSRL on a ADSC Quantum-315 CCD detector and processed with the HKL package²³ (see Table 1). PA was positioned in the unit cell by Molecular Replacement (Protein Data Bank (PDB) ID code 1acc)⁶ using MOLREP, and refined with REFMAC version 5.0 (ref. 24). Density for the MIDAS Mn²⁺ ion and upper loops of the receptor was evident in this map, and a molecule

of CMG2 (PDB ID code 1SHT)¹¹ was manually placed in the electron density. Model building was performed with O²⁵ and TURBOFRODO (A. Roussel and C. Cambillau, Silicon Graphics), and the solvent structure was built with ARP/wARP 6.0 (ref. 26). Although the random errors in the diffraction data are high, owing to the small crystal size, the final refinement statistics and maps are excellent (Table 1). Thus, the final *R*-factors are *R*_{free} = 26.6% and *R*_{work} = 20.7% overall, and *R*_{free} = 37.2% and *R*_{work} = 27.5% in the outer resolution bin, with root-mean-square deviations (r.m.s.d.) from ideal values of 0.017 Å for bond lengths and 1.65° for angles. Stereochemistry is excellent as assessed with PROCHECK²⁴, and the model is consistent with composite simulated annealing omit maps (3,000 °C) calculated in CNS²⁷. The model comprises residues 16–735 of PA; 41–210 of CMG2, with the exception of three loops (residues 159–174, 276–287 and 304–319) in PA for which no electron density was observed; 139 water molecules; two Ca²⁺ ions in PA domain I; two Na⁺ ions; one PEG molecule; and one Mn²⁺ ion at the MIDAS site. The *B* factors for the Ca²⁺ and Mn²⁺ ions (27–33 Å²) are higher than for the coordinating residues (16–20 Å²). Although the MIDAS metal ion *in vivo* is likely to be Mg²⁺, we have previously shown for integrin I domains that the stereochemistry of the open conformation is not dependent on the nature of the metal ion⁵. The bond lengths to the Mn²⁺ ion are 2.1 ± 0.2 Å, identical to those observed in integrin–ligand complexes^{5,13,28}. PA domain I (residues 16–258) undergoes a small rotation as a consequence of crystal constraints when compared with the structure of isolated PA such that the r.m.s.d. values for the superposition of the two molecules are 1.44, 0.58 and 0.79 Å for residues 16–735, 259–735 and 16–258 respectively. CMG2 residues 41–200 superimpose with a r.m.s.d. of 0.60 with the isolated protein¹¹, while the C-terminal helix (residues 201–210) shifts downwards by one helical turn.

Received 6 May; accepted 18 June 2004; doi:10.1038/nature02763.
Published online 4 July 2004.

- Moayeri, M. & Leppla, S. H. The roles of anthrax toxin in pathogenesis. *Curr. Opin. Microbiol.* **7**, 19–24 (2004).
- Abrami, L., Liu, S., Cosson, P., Leppla, S. H. & Vander Goot, F. G. Anthrax toxin triggers endocytosis of its receptor via a lipid raft-mediated clathrin-dependent process. *J. Cell Biol.* **160**, 321–328 (2003).
- Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414**, 225–229 (2001).
- Scobie, H. M., Rainey, G. J., Bradley, K. A. & Young, J. A. Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor. *Proc. Natl Acad. Sci. USA* **100**, 5170–5174 (2003).
- Emsley, J., Knight, C. G., Farnedale, R. W., Barnes, M. J. & Liddington, R. C. Structural basis of collagen recognition by integrin α2β1. *Cell* **101**, 47–56 (2000).
- Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385**, 833–838 (1997).
- Benson, E. L., Huynh, P. D., Finkelstein, A. & Collier, R. J. Identification of residues lining the anthrax protective antigen channel. *Biochemistry* **37**, 3941–3948 (1998).
- Nassi, S., Collier, R. J. & Finkelstein, A. PA63 channel of anthrax toxin: an extended β-barrel. *Biochemistry* **41**, 1445–1450 (2002).
- Frankel, A. E., Koo, H.-K., Leppla, S. H., Duesbury, N. S. & Vande Woude, G. F. Novel protein targeted therapy of metastatic melanoma. *Curr. Pharm. Des.* **9**, 2060–2066 (2003).
- Lee, J.-O., Rieu, P., Arnaout, M. A. & Liddington, R. C. Crystal structure of the A-domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* **80**, 631–635 (1995).
- Lacy, D. B., Wigelsworth, D. J., Scobie, H. M., Young, J. A. & Collier, R. J. Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: An anthrax toxin receptor. *Proc. Natl Acad. Sci. USA* **101**, 6367–6372 (2004).
- Wigelsworth, D. J. *et al.* Binding stoichiometry and kinetics of the interaction of a human anthrax toxin receptor, CMG2, with protective antigen. *J. Biol. Chem.* **279**, 23349–23356 (2004).
- Shimaoka, M. *et al.* Structures of the αI domain and its complex with ICAM-1 reveal a shape-shifting pathway for integrin regulation. *Cell* **112**, 99–111 (2003).
- Rosovitz, M. J. *et al.* Alanine scanning mutations in domain 4 of anthrax toxin protective antigen reveal residues important for binding to the cellular receptor and to a neutralizing monoclonal antibody. *J. Biol. Chem.* **278**, 30936–30944 (2003).
- Bradley, K. A. *et al.* Binding of anthrax toxin to its receptor is similar to α integrin–ligand interactions. *J. Biol. Chem.* **278**, 49342–49347 (2003).
- Mourez, M. *et al.* Mapping dominant-negative mutations of anthrax protective antigen by scanning mutagenesis. *Proc. Natl Acad. Sci. USA* **100**, 13803–13808 (2003).
- Miller, C. J., Elliott, J. L. & Collier, R. J. Anthrax protective antigen: prepore-to-pore conversion. *Biochemistry* **38**, 10432–10441 (1999).
- Petosa, C. in *Crystal Structure of the Anthrax Protective Antigen*. Thesis, Harvard Univ (1995).
- Song, L. *et al.* Structure of staphylococcal α-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1866 (1996).
- Nanda, A. & St Croix, B. Tumor endothelial markers: new targets for cancer therapy. *Curr. Opin. Oncol.* **16**, 44–49 (2004).
- Nanda, A. *et al.* TEM8 interacts with the cleaved C5 domain of collagen alpha 3(VI). *Cancer Res.* **64**, 817–820 (2004).
- Liu, S., Schubert, R. L., Bugge, T. H. & Leppla, S. H. Anthrax toxin: structures, functions and tumour targeting. *Expert Opin. Biol. Ther.* **3**, 843–853 (2003).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project, No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models into electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

- Lee, J.-O., Bankston, L. A., Arnaout, M. A. & Liddington, R. C. Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure* **3**, 1333–1340 (1995).
- Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).

Acknowledgements We thank the NIH and the DOD for financial support, and the DOE and staff at the SSRL for synchrotron access and support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.C.L. (rlidding@burnham.org). The atomic coordinates have been deposited in the Protein Data Bank under accession code 1T6B.

Cell cycle regulation of central spindle assembly

Masanori Mishima¹, Visnja Pavicic¹, Ulrike Grüneberg², Erich A. Nigg² & Michael Glotzer¹

¹Research Institute of Molecular Pathology, Dr. Bohrgasse 7, A-1030 Vienna, Austria

²Max-Planck-Institute für Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany

The bipolar mitotic spindle is responsible for segregating sister chromatids at anaphase. Microtubule motor proteins generate spindle bipolarity and enable the spindle to perform mechanical work¹. A major change in spindle architecture occurs at anaphase onset when central spindle assembly begins. This structure regulates the initiation of cytokinesis and is essential for its completion². Central spindle assembly requires the centralspindlin complex composed of the *Caenorhabditis elegans* ZEN-4 (mammalian orthologue MKLP1) kinesin-like protein and the Rho family GAP CYK-4 (MgcRacGAP). Here we describe a regulatory mechanism that controls the timing of central spindle assembly. The mitotic kinase Cdk1/cyclin B phosphorylates the motor domain of ZEN-4 on a conserved site within a basic amino-terminal extension characteristic of the MKLP1 subfamily. Phosphorylation by Cdk1 diminishes the motor activity of ZEN-4 by reducing its affinity for microtubules. Preventing Cdk1 phosphorylation of ZEN-4/MKLP1 causes enhanced metaphase spindle localization and defects in chromosome segregation. Thus, phosphoregulation of the motor domain of MKLP1 kinesin ensures that central spindle assembly occurs at the appropriate time in the cell cycle and maintains genomic stability.

At the metaphase–anaphase transition, the anaphase-promoting complex triggers proteolysis of cyclin B (an activating subunit of the mitotic kinase Cdk1) and sister chromatid separation. Chromosomes move polewards and non-kinetochore spindle microtubules become bundled, initiating assembly of the central spindle, a structure that has important roles in cytokinesis. In *C. elegans* embryos and other animal cells, central spindle assembly requires centralspindlin³. Many proteins that regulate mitosis and cytokinesis re-localize upon anaphase onset. For example, Aurora B and its associated subunits dissociate from centromeres and concentrate on the central spindle^{4–6}. Similarly, anaphase onset triggers redistribution of centralspindlin (Fig. 1a, b). In metaphase, centralspindlin is diffuse and in anaphase it localizes to the microtubules positioned between the separating chromosomes, as seen previously^{7–10}. ZEN-4 (also known as CeMKLP1) colocalizes with the proline-directed phosphatase CDC-14 (ref. 11) and depletion of CDC-14 prevents ZEN-4 localization¹². Non-degradable cyclins stabilize Cdk1 activity and prevent central spindle assembly^{13,14}. Together these data

indicate that ZEN-4 localization is controlled, directly or indirectly, by a proline-directed kinase-phosphatase switch. However, it is not yet known in molecular detail how ZEN-4 localization is controlled and whether this regulatory pathway has important biological implications.

As a first step in evaluating how centralspindlin localization is regulated, we assessed whether the motor domain of ZEN-4 (amino acids 1–434; referred to hereafter as ZEN-4^{MOT}) is a substrate for Cdk1/cyclin B *in vitro*. ZEN-4 and its orthologues contain an N-terminal Cdk1 consensus site (S/T)-P-X-(K/R) that is conserved among most animal species (Fig. 1c). ZEN-4^{MOT} was efficiently phosphorylated by Cdk1/cyclin B. In contrast, a protein variant in which the conserved phosphoacceptor site was substituted by alanine (ZEN-4^{MOT-T9A}) was not phosphorylated (Fig. 1d). The motor domain of the mammalian orthologue of ZEN-4, MKLP1, was also phosphorylated by Cdk1/cyclin B on the corresponding site, T8, and one additional site, T450 (Fig. 1e). The phosphorylation site in ZEN-4, T9, is located in a ~20-residue extension N-terminal to the catalytic core that is found in the MKLP1 subfamily and in some, but not all, kinesin-like proteins. In the MKLP1 subfamily this extension is highly basic; phosphorylation at this site would reduce the overall charge on this extension.

MKLP1 and its orthologues have not previously been shown to support microtubule motility in standard gliding assays. However, lysates from bacteria expressing MKLP1 have been shown to support antiparallel microtubule gliding at $0.07 \mu\text{m s}^{-1}$ (ref. 15). Therefore, we adapted the conventional gliding assay, using purified ZEN-4 motor domain biotinylated at its carboxy terminus to facilitate oriented adsorption to a glass surface¹⁶. ZEN-4^{MOT} supported robust microtubule motility *in vitro* over a range of ionic conditions (25–250 mM KCl), at an average velocity of $0.16 \mu\text{m s}^{-1}$ (Supplementary Movie 1). ZEN-4 constructs that contain the entire neck region translocated microtubules more rapidly (average velocity $0.4 \mu\text{m s}^{-1}$) (unpublished results, M.M. and M.G.). We note that microtubule gliding does not require the ZEN-4 binding partner CYK-4.

To assess the function of the N-terminal extension that contains the Cdk1/cyclin B phosphorylation site, we compared the ATPase activity and motility of ZEN-4^{MOT} with ZEN-4^{ΔN} (amino acids 21–434). Under physiological salt conditions (150 mM KCl), the microtubule-activated ATPase activity of ZEN-4^{ΔN} is reduced to near background levels (Fig. 2a). We also could not observe gliding of microtubules by ZEN-4^{ΔN} at physiological salt concentrations; ATP perfusion caused release of most microtubules. However, at 25 mM KCl, ATP caused partial release of microtubules (Fig. 2b) and some of the remaining microtubules moved briefly with a velocity similar to that of ZEN-4^{MOT}. These data indicate that the N-terminal extension is critical for motility under physiological ionic conditions. Because microtubules bound to ZEN-4^{ΔN} in the absence of ATP and dissociated in the presence of ATP, the basic N terminus seems to stabilize the interaction between ZEN-4 and microtubules during the weak binding state when ADP (or ADP + Pi) is bound, analogous to the function of the K-loop in the KIF1A motor¹⁷. Moreover, in conventional kinesin the N terminus is adjacent to loop 12 (Supplementary Movie 2), the site of the K-loop insertion, which binds to the negatively charged C-terminal tail of tubulin. By analogy, the N terminus of ZEN-4 probably interacts with the C terminus of tubulin.

Because the N-terminal extension has a critical role in the motor activity of ZEN-4, phosphorylation of T9 might regulate this motor. Indeed, phosphorylated ZEN-4^{MOT} had greatly reduced activity in the ATPase assay (Fig. 2c). We also assessed the ability of phosphorylated ZEN-4^{MOT} to support microtubule gliding. Whereas ZEN-4^{MOT} efficiently bound microtubules in the presence of ATP and supported microtubule gliding at high motor density (5×10^4 motors μm^{-2} , 150 mM KCl, average velocity $0.14 \mu\text{m s}^{-1}$), after phosphorylation of T9 fewer microtubules remained bound to the

surface (Fig. 2d). The remaining microtubules moved with the same average velocity as observed with the unphosphorylated protein, although the gliding was saltatory. At lower motor density (10^4 motors μm^{-2}), unphosphorylated ZEN-4^{MOT} still supported robust motility, but phosphorylated ZEN-4^{MOT} did not bind microtubules

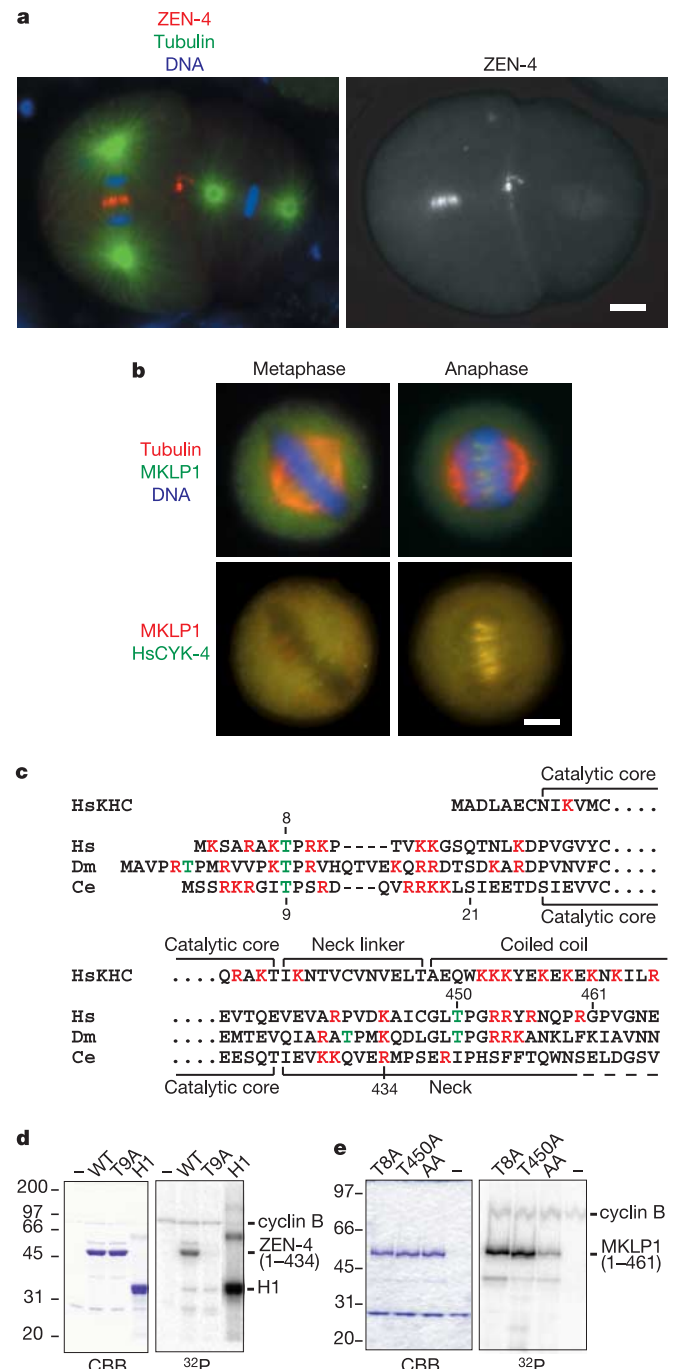


Figure 1 ZEN-4/MKLP1 localizes upon anaphase onset and is a Cdk1 substrate *in vitro*. **a**, **b**, In metaphase, ZEN-4/MKLP1 weakly associates with the spindle and concentrates on the central spindle upon anaphase onset in *C. elegans* embryos (**a**) and HeLa cells (**b**). **c**, ZEN-4/MKLP1 has a conserved consensus site for phosphorylation by Cdk1 N-terminal to the kinesin catalytic core. HsKHC, human kinesin heavy chain; Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*. **d**, ZEN-4^{MOT} is phosphorylated by Cdk1/cyclin B *in vitro*. ZEN-4^{MOT-T9A} is not phosphorylated (H1, histone H1). WT, wild type. CBB, coomassie blue. **e**, MKLP1 is phosphorylated by Cdk1/cyclin B on T8 and T450. MKLP1 (1–461) mutants T8A, T450A and T8A/T450A, were incubated with Cdk1/cyclin B *in vitro*. Scale bars, 5 μm .

at all (Fig. 2d). We quantified the frequency of microtubules detaching from the surface and found that T9 phosphorylation destabilizes the interaction between ZEN-4^{MOT} and microtubules (Fig. 2e). Cdk1/cyclin B had no effect on ZEN-4^{MOT-T9A}, indicating that these effects were due to phosphorylation of T9 (Fig. 2d, e). Phosphorylation of ZEN-4^{MOT} did not significantly alter the estimated maximum ATPase rate but greatly increased the concentration of microtubules required for maximal activity (Fig. 2f).

Phosphorylation of the ZEN-4 N-terminal extension regulates the biochemical properties of ZEN-4 *in vitro*, therefore we evaluated whether these sites are phosphorylated *in vivo*. Because it is not feasible to obtain mitotic *C. elegans* embryos for biochemical analysis, we investigated the *in vivo* phosphorylation state of the human orthologue MKLP1 by two-dimensional phosphopeptide mapping. The motor domain of MKLP1 contains two Cdk1/cyclin B consensus sites, T8 and T450, that were efficiently phosphorylated by Cdk1/cyclin B (Fig. 1e). Phosphopeptide maps of recombinant MKLP1 motor domain (1–461) with T8A or T450A substitutions revealed one major phosphopeptide in each case, that were used as migration standards (Fig. 3b). HeLa cells were arrested in mitosis, labelled with ³²P orthophosphate and MKLP1 was immunopurified. MKLP1 was phosphorylated, as was the coprecipitating HsCYK-4/MgcRacGAP (Fig. 3a). On phosphopeptide maps, MKLP1 from ³²P-labelled mitotic cells gave six major phospho-

peptides. Two phosphopeptides co-migrated with the phosphoT8 and phosphoT450 standards (Fig. 3b). Thus, both T8 and T450 of MKLP1 are phosphorylated in mitotic cells.

If MKLP1 is delocalized during metaphase as a consequence of phosphorylation by Cdk1/cyclin B, then these sites should be dephosphorylated during anaphase to allow MKLP1 to localize. To evaluate the changes in phosphorylation of the Cdk1/cyclin B sites during the exit from mitosis and the onset of cytokinesis we used an anti-phosphothreonine-proline (TP) antibody on immunoblots of MKLP1 immunoprecipitates. This antibody primarily detects phosphorylation of MKLP1 on T8 and T450 because its reactivity is greatly diminished when these residues are replaced by serines (Fig. 3c). Endogenous MKLP1, immunoprecipitated from nocodazole-arrested cells, was phosphorylated (Fig. 3d; 0 min) and this phosphorylation was stable for ~25 min after release from mitotic arrest, after which time the phosphorylation level was significantly reduced. To show that this phosphorylation is dependent on Cdk1 *in vivo*, mitotic cells were treated with the Cdk1 inhibitor roscovitine¹⁸. This compound induced rapid dephosphorylation of endogenous MKLP1 on TP sites (Fig. 3e) and caused the protein to concentrate on spindles (Fig. 3f). We conclude that MKLP1 is phosphorylated by Cdk1/cyclin B, and this kinase regulates the localization of MKLP1 *in vivo*.

A candidate phosphatase that could remove this inhibitory

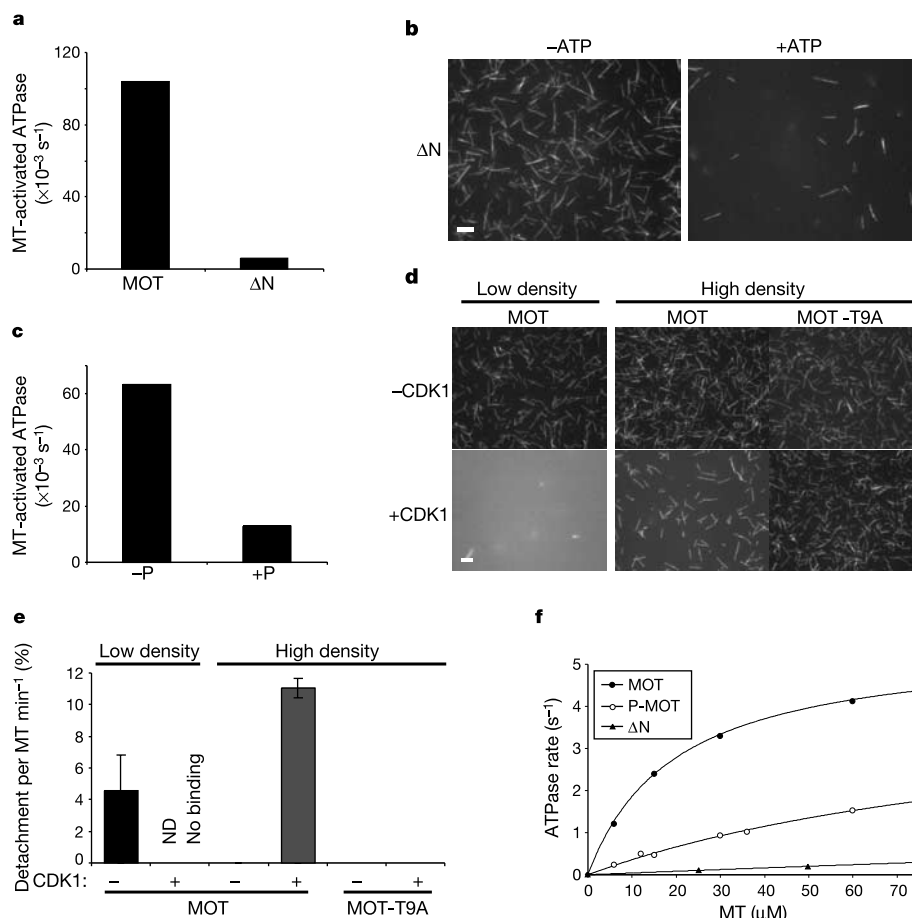


Figure 2 Microtubule (MT) binding by ZEN-4 is inhibited by Cdk1/cyclin B phosphorylation. **a**, ATPase activity of ZEN-4^{MOT} and ZEN-4 ^{ΔN} . **b**, Microtubule binding by ZEN-4 ^{ΔN} in the presence and absence of ATP. **c**, ATPase activity of ZEN-4^{MOT} with (+P) and without (–P) stoichiometric phosphorylation by Cdk1/cyclin B. **d**, Microtubule capture by ZEN-4^{MOT} and ZEN-4^{MOT-T9A} in the presence or absence of Cdk1/cyclin B phosphorylation at two densities of ZEN-4. Images shown are frames from a motility assay after ATP addition. **e**, Quantification of the detachment probability ($\pm$ s.e.m.) at two

densities of ZEN-4 as in **d**, **f**. Kinetics of ATP hydrolysis by ZEN-4^{MOT}, phospho-ZEN-4^{MOT} and ZEN-4 ^{ΔN} as a function of microtubule concentration. For MOT and P-MOT, the curves shown represent Michaelis-Menten equations with the following constants (maximal ATPase rates (s^{-1}): 5.5, 4.4 and $K_{0.5, \text{MT}}$ (μM): 19.6 to 112 for MOT and P-MOT, respectively). Assays shown in **a–f** were performed at 150 mM KCl and the assays in **f** were performed at 75 mM KCl. Scale bars, 5 μm .

Table 1 Impaired rescue of *zen-4(w35)* by *zen-4* T9A

Line	Embryos	Viable adults (%)	Rol animals	Bli animals	Array transmission (%)	Rescue efficiency (%)
<i>zen-4</i>						
1	1,178	46.2	281	66	45.0	49.8
2	789	30.8	101	31	33.0	47.6
3	993	44.1	203	48	39.7	48.7
4	1,491	47.6	355	88	42.9	55.0
5	800	44.9	204	69	46.6	74.1
Total	5,251	43.7	1,144	302	42.3	54.4
<i>zen-4</i> T9A						
1	1,355	41.3	287	27	48.9	16.3
2	1,157	46.2	270	11	49.5	7.7
3	803	34.1	125	0	45.6	0.0
4	2,010	48.2	361	0	37.3	0.0
Total	5,325	43.8	1,043	38	43.8	6.5

Extrachromosomal transgenic arrays marked by *rol-6(su1006)* were generated in *zen-4(w35)* *bli-6(sc16)* *unc-44(e1260)* *lag-1(q385)*. Embryos were collected from hermaphrodites and scored for fully developed adults, Rol animals, and Bli animals (all of which were Rol). Rescue efficiency = number of Bli animals/[(number of embryos × array transmission %)/4], where array transmission % = (number of Rol-number of Bli animals)/(number of viable adults-number of Bli animals). No Bli animals were observed in three transgenic lines generated with *rol-6(su1006)* and without *zen-4* (>1,000 embryos each). Results shown are the sum of two independent experiments.

phosphorylation and allow MKLP1-dependent central spindle assembly is CDC14. Therefore, we tested whether recombinant human CDC14A could dephosphorylate MKLP1 phosphorylated by Cdk1/cyclin B. Phospho-MKLP1 fragments MKLP1 T8A and MKLP1 T450A, in which T450 and T8 are the major phosphorylation sites, respectively, were dephosphorylated by hCDC14A (Fig. 3g). This phosphatase activity was sensitive to orthovanadate, and human CDC14A with a mutation in the catalytic cysteine (PD) was inactive. *C. elegans* CDC-14 also dephosphorylated ZEN-4 (data not shown). The function of CDC-14 *in vivo* is consistent with it activating ZEN-4 upon anaphase onset.

Phosphorylation of ZEN-4 at T9 significantly affects its motor activity *in vitro* and the corresponding site is phosphorylated in human cells. To determine whether this phosphorylation is essential *in vivo*, we generated a *C. elegans* strain containing a null allele (*w35*) of ZEN-4 linked to a recessive visible marker, *bli-6*, which causes cuticle blistering. The *w35* allele causes embryonic lethality¹⁰; this lethality is associated with cytokinesis defects in embryos at the 100–300-cell stage (Fig. 4a). We generated multiple independent transgenic lines expressing either ZEN-4 or ZEN-4 T9A. Whereas wild-type ZEN-4 could effectively rescue *zen-4(w35)*, the T9A substitution reduced the rescue efficiency ~8-fold (Table 1). Thus, ZEN-4 with a non-phosphorylatable residue at position 9 is defective *in vivo*.

To determine whether the phosphorylation of MKLP1 on T8 and T450 regulates its localization in human cells, we expressed Myc-tagged wild-type and mutant proteins (where T8 and T450 were mutated to alanine; MKLP1-AA) in HeLa cells. Although ectopically expressed wild-type protein localized on metaphase spindles to a greater extent than the endogenous protein, MKLP1-AA was significantly more enriched on the spindle (Fig. 4b), suggesting that phosphorylation regulates MKLP1 localization.

Phosphorylation of MKLP1 during metaphase may ensure that central spindle assembly does not occur before anaphase onset. For example, if kinetochore microtubules become bundled to anti-parallel microtubules from the opposite spindle pole during metaphase, efficient segregation of chromosomes during anaphase could be impaired. To test this hypothesis, synchronized HeLa cells expressing MKLP1 or MKLP1-AA were fixed at anaphase and the frequency of aberrant chromosome segregation was assessed. An increase in chromosome bridges was observed with both MKLP1 derivatives. However, the frequency of lagging chromosomes was significantly higher when cells expressed MKLP1-AA compared with MKLP1 (Fig. 4c, d). Therefore, phosphorylation of MKLP1 during metaphase inhibits its association with the spindle and prevents the induction of aneuploidy.

To determine whether CDC-14 is the critical phosphatase that

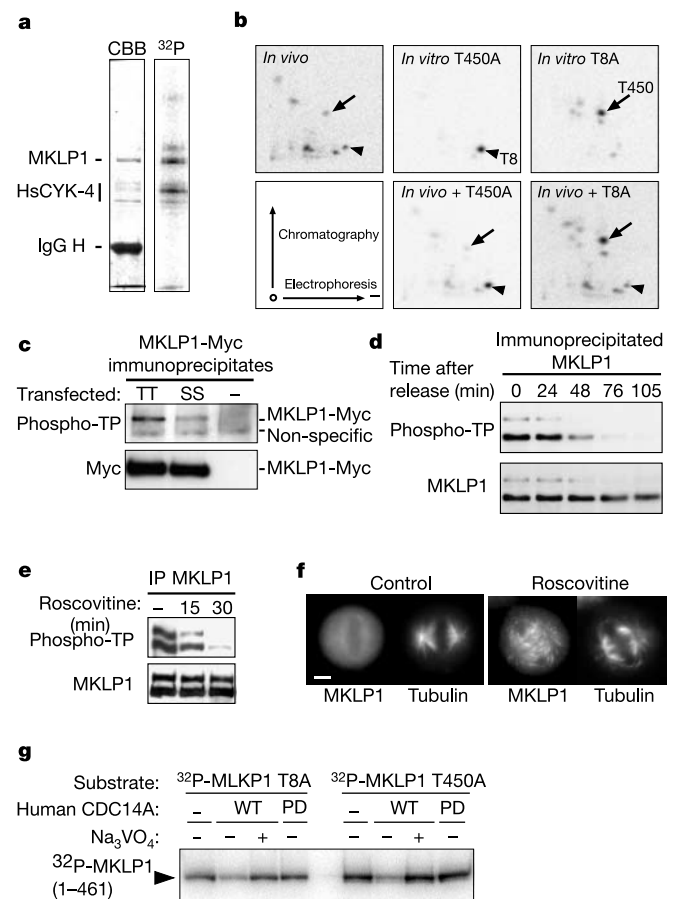


Figure 3 MKLP1 is phosphorylated on Cdk1/cyclin B sites in metaphase and dephosphorylated during anaphase. **a**, Immunoprecipitation of centralspindlin from metabolically labelled mitotic HeLa cells. **b**, Two-dimensional peptide maps of phosphoMKLP1 and *in vitro* generated phosphoMKLP1 T450A and phosphoMKLP1 T8A. **c**, Immunoprecipitates of MKLP1-Myc (TT) and MKLP1 T8ST450S-Myc (SS) were immunoblotted with a phosphoTP antibody revealing that this antibody primarily reacts with phosphoT8 and phosphoT450. **d**, Immunoblot with phosphoTP antibody of MKLP1 immunoprecipitated from cells released from a nocodazole block for the indicated times. **e**, Immunoblot with phosphoTP antibody of MKLP1 immunoprecipitated from cells arrested in mitosis with nocodazole and treated with roscovitine for the indicated times. **f**, Localization of MKLP1 and tubulin in cells treated as in **e**, except that 20 ng ml⁻¹ nocodazole was used to induce arrest. **g**, Dephosphorylation of phosphoMKLP1 T450A and phosphoMKLP1 T8A by human CDC14A (WT) *in vitro*. Mutation of the catalytic cysteine 278 to serine (PD) or addition of sodium vanadate renders the enzyme inactive. Scale bars, 5 μ m.

dephosphorylates T9, we investigated whether, unlike wild-type ZEN-4, the T9A mutant can localize during anaphase in embryos depleted of CDC-14. *C. elegans* strains, expressing ZEN-4::GFP (green fluorescent protein) or ZEN-4::GFP T9A in the germline, were generated. CDC-14 was depleted by RNAi and the behaviour of ZEN-4::GFP was monitored by time-lapse microscopy. ZEN-4::GFP rarely localized to the central spindle when CDC-14 was depleted (3 of 19 embryos), as shown previously¹². In contrast, the non-phosphorylatable mutant ZEN-4::GFP T9A localized in nearly all *cdc-14(RNAi)* embryos (20 of 22 embryos) (Fig. 4e). We conclude that localization of ZEN-4::GFP T9A requires less CDC-14 activity than does localization of ZEN-4::GFP. As a control, we found that strains expressing wild-type or T9A derivatives of ZEN-4::GFP were equally sensitive to *rhoA(RNAi)* and *let-21(RNAi)* (Fig. 4e and data not shown). Thus, dephosphorylation of phospho-T9 is CDC-14-dependent *in vivo* and may well be direct. Lines expressing ZEN-4 T9A segregated males at elevated frequency compared with lines expressing the wild-type protein (ZEN-4, 0.16%, $n = 3,703$; ZEN-4::T9A, 0.55%, $n = 4,883$). In *C. elegans*, males arise at low frequency owing to non-disjunction of the X chromosome in the hermaphrodite germ line, hence these data provide further evidence

that failure to phosphorylate this site induces defects in chromosome segregation.

We have shown that ZEN-4/CeMKLP1 has microtubule motility *in vitro*. Motor activity seems to be essential *in vivo* as well, because mutations in the ATP-binding P-loop prevent discrete localization of MKLP1 to the central spindle and prevent normal central spindle assembly and completion of cytokinesis^{19,20}. Microtubule binding by MKLP1 involves a highly basic N-terminal extension of the motor, which may interact with the C-terminal acidic tail of tubulin in a manner similar to the K-loop in KIF1A (ref. 21). The function of this basic region is regulated by direct phosphorylation by Cdk1/cyclin B *in vitro*, and the equivalent site in MKLP1 was shown to be phosphorylated *in vivo*. Basic clusters are present in the N terminus of numerous kinesin-like proteins, and many of these clusters contain, or are adjacent to, phosphoacceptor sites. The microtubule depolymerizing activity of MCAK has been shown to be inhibited by Aurora B phosphorylation of a serine residue within a region rich in basic residues^{22–24}. Phosphorylation of residues embedded in basic stretches that may interact with the C-terminal acidic tail of tubulin might therefore be a general mechanism to regulate kinesin-like proteins.

ZEN-4 mediates central spindle assembly by bundling the plus ends of antiparallel microtubules. This activity may need to be inhibited during metaphase to prevent bundling of a kinetochore fibre to microtubules nucleated by the other spindle pole. In budding yeast, Cdk1/cyclin B and Cdc14p antagonistically control the localization of Aurora B-Incenp (inner centromere protein complex) and chromosome segregation defects result from the precocious localization of this complex⁵. PRC1, a microtubule binding protein important for central spindle assembly, is inhibited by phosphorylation on Cdk1/cyclin B consensus sites²⁵. Thus, during metaphase, multiple mechanisms prevent microtubule bundling and thereby maintain a stable karyotype. □

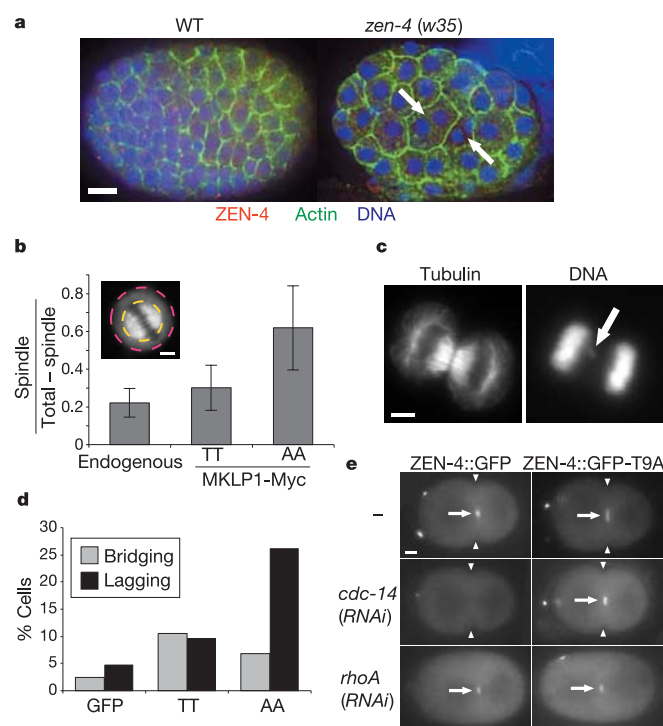


Figure 4 Non-phosphorylatable MKLP1 induces lagging chromosomes, and non-phosphorylatable ZEN-4 localizes in embryos depleted of CDC-14. **a**, *zen-4(w35)* embryos are cytokinesis-defective during early embryogenesis (arrows, binucleate cells) (ZEN-4, red; DNA, blue; actin, green). Homozygous mutant embryos were identified by a strong reduction in the ZEN-4 staining at equivalent stages and this was always associated with multinucleated blastomeres. **b**, Quantification of the spindle localization of MKLP1-Myc, MKLP1-AA-Myc and endogenous MKLP1. The ratio ($\pm$ s.d.) of the spindle fluorescence to non-spindle fluorescence was calculated as described in the methods. The inset shows MKLP1AA-Myc concentrates on the metaphase spindle in HeLa cells. Circles indicate regions used for the quantification of the extent of spindle localization. **c**, Overexpression of MKLP1-AA causes lagging chromosomes (arrow). **d**, Quantification of the frequency of anaphase bridging and lagging chromosomes in cells transfected with MKLP1 ($n = 105$ cells), MKLP1-AA ($n = 109$) or GFP ($n = 84$). **e**, ZEN-4 T9A, but not ZEN-4, can localize in *cdc-14(RNAi)* embryos (arrows), whereas both strains were equally susceptible to *rhoA(RNAi)* as evidenced by the absence of cleavage furrow ingression (arrowheads). Scale bars, 5 μ m.

Methods

Nematode strains and alleles

The following strains and alleles were used: EH135 (*unc-44(e362) bli-6(sc16) IV*), SU62 (*unc-44(e1260) lag-1(q385)/zen-4(w35) IV*), MG376 (*zen-4(w35) bli-6(sc16)/unc-44(e1260) lag-1(q385)*). Some strains used in this study were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH NCRR. For rescue assays, simple transgene arrays were prepared as described²⁶ using pRF4 at 80 ng μ l⁻¹, and MP66 (*zen-4*) and MM196 (*zen-4 T9A*) at 10 ng μ l⁻¹. To analyse ZEN-4::GFP localization in the early embryo, complex arrays that rescue *zen-4(or153ts)* were prepared as described²⁷. Depletion of CDC-14 by RNAi was performed as described¹², using dsRNA corresponding to nucleotides 51–1214 of C17G10.4C.

Cell culture and transfection

HeLa cells were cultured in DMEM with fetal bovine serum. MKLP1 derivatives, containing a C-terminal triple Myc tag, were transfected using LipofectAmine PLUS (Invitrogen). To analyse chromosome segregation defects, cells were synchronized with a double thymidine block and transfected with Eugene 6 (Roche) 23 h before release from the second block. Eleven hours after release of the block, mitotic cells were collected by shake-off, and re-plated and fixed 30 min later. For inhibitor experiments, cells were synchronized in mitosis with 40 ng ml⁻¹ nocodazole, then 100 μ M roscovitine (Calbiochem) was added for the indicated times before analysis.

Protein preparation

For kinase assays, ZEN-4 and MKLP1 motor domains were cloned into pCBD-TEV and expressed in bacteria as fusion proteins with chitin-binding domain at the N terminus³. For gliding assays these constructs were modified to include a C-terminal flexible linker and biotin-acceptor sequence (GTGSG-GLNDIFEAAQKIEWHE) (ref. 28). Recombinant proteins were affinity-purified with chitin beads and eluted with TEV (tobacco etch virus) protease. All the proteins except those used in Fig. 1d were further purified by ion exchange chromatography (MonoS for ZEN-4^{MO} (1–434) and MKLP1 constructs, and MonoQ for ZEN-4^{AN} (21–434)) followed by gel filtration (Superdex 200, Pharmacia) chromatography. Human CDC14A was amplified from human testis complementary DNA and cloned into pGEX-5X-1. Phosphatase-dead human CDC14A was generated by mutating cysteine 278 to serine. The resulting GST fusion proteins were expressed and purified according to standard procedures.

Kinase and phosphatase assays

For kinase assays, 0.1–0.2 mg ml⁻¹ ZEN-4 and MKLP1 proteins were incubated with purified recombinant human Cdk1/GST-cyclin B in kinase buffer (50 mM KCl, 2 mM

MgCl₂, 1 mM EGTA, 10 mM PIPES (pH 6.8), 1 mM DTT, 50 μ M ATP and 50 μ Ci ml⁻¹ [³²P]ATP at 25 °C (ref. 29). To prepare substrates for the phosphatase assay, free radiolabelled ATP was removed by repeated dilution and concentration with microcon-30 filtration devices. For phosphatase assays, 0.6 mg ml⁻¹ phospho-MKLP1 was incubated with 60 μ M ml⁻¹ bacterially expressed recombinant GST-human CDC14A in phosphatase buffer (50 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 1 mM DTT and 10 mM PIPES (pH 6.8)) at 25 °C.

ATPase and microtubule motility assays

For the ATPase assay in Fig. 2b, phosphorylated ZEN-4 motor domain was purified from unphosphorylated ZEN-4 by Mono S and Superdex 200 (Pharmacia) chromatography. The control unphosphorylated sample was also re-purified with these columns. ZEN-4 motor domains were incubated with 5 μ M taxol-stabilized microtubules from phosphocellulose-purified bovine brain tubulin in ATPase buffer (150 mM KCl, 10 mM PIPES (pH 6.8), 2 mM MgCl₂, 1 mM EGTA, 2 mM ATP, 1 mM DTT and 5 μ M taxol) at 25 °C. In Fig. 2f, 75 mM KCl was used in the ATPase buffer. Free phosphate release was measured by colorimetry with molybdate/malachite green. For surface binding and gliding assays, bacterially biotinylated ZEN-4 motor domains were immobilized on a coverslip sequentially coated with biotin-BSA and NeutrAvidin (Molecular Probes)¹⁶. The extent of biotinylation was estimated by western blotting with an anti-biotin antibody (Sigma).

Analysis of *in vivo* phosphorylation state

For mitotic synchronization, cells were arrested for 14 h with 2.5 mM thymidine. The cells were released for 7.5 h and then 50 ng ml⁻¹ nocodazole was added for 3.5 h to establish an M-phase arrest. For phosphopeptide mapping, 0.3 mCi ml⁻¹ ³²P-orthophosphate was added with the nocodazole. Lysates were prepared with IP buffer (150 mM NaCl, 10 mM NaF, 40 mM β -glycerophosphate, 20 mM HEPES (pH 7.5), 2 mM MgCl₂, 10 mM EDTA, 0.5% Triton X-100, 1 mM DTT, 1 mM PMSF, 50 μ g ml⁻¹ leupeptin, 50 μ g ml⁻¹ pepstatin and 1 μ M microcystin). MKLP1 was immunoprecipitated with anti-MKLP1 antibody bound to protein-A Sephadex beads and resolved by SDS-polyacrylamide gel electrophoresis. Phosphoproteins were digested in-gel with trypsin. Eluted peptides were analysed using pH 1.9 electrophoresis buffer³⁰. Phosphorylation at TP sites was detected with an anti-phosphoTP monoclonal antibody (Cell Signalling Technology).

Immunolocalization and microscopy

HeLa cells were fixed with -20 °C methanol and stained as described before. To detect the Myc-epitope, an affinity purified rabbit antiserum (Gramsch Laboratories) was used. To estimate the accumulation of Myc-tagged MKLP1, the area of the cell (*a_T*) and of the spindle region (*a_S*), and the mean intensity of the cell (*m_T*) and of the spindle region (*m_S*) were measured from unprocessed 12-bit images with ImageJ (<http://rsb.info.nih.gov/ij/>). The ratio between spindle-derived signal and the non-spindle-derived signal was calculated as (*m_S-m_T*)/(*a_S*(*m_T-a_T*)/*a_S*). Images were collected on either a Zeiss Axioplan II imaging microscope fitted with a CoolSnap Fx CCD camera or a DeltaVision Imaging System (Applied Precision).

Received 27 April; accepted 16 June 2004; doi:10.1038/nature02767.

Published online 28 July 2004.

- Scholey, J. M., Brust-Mascher, I. & Mogilner, A. Cell division. *Nature* **422**, 746–752 (2003).
- Glötzer, M. Animal cell cytokinesis. *Annu. Rev. Cell Dev. Biol.* **17**, 351–386 (2001).
- Mishima, M., Kaitna, S. & Glötzer, M. Central spindle assembly and cytokinesis require a kinesin-like protein/RhoGAP complex with microtubule bundling activity. *Dev. Cell* **2**, 41–54 (2002).
- Adams, R. R., Carmona, M. & Earnshaw, W. C. Chromosomal passengers and the (aurora) ABCs of mitosis. *Trends Cell Biol.* **11**, 49–54 (2001).
- Pereira, G. & Schiebel, E. Separase regulates INCENP-Aurora B anaphase spindle function through Cdc14. *Science* **302**, 2120–2124 (2003).
- Romano, A. *et al.* CSC-1: A subunit of the Aurora B kinase complex that binds to the Survivin-like protein BIR-1 and the Incenp-like protein ICP-1. *J. Cell Biol.* **161**, 229–236 (2003).
- Sellitto, C. & Kuriyama, R. Distribution of a matrix component of the midbody during the cell cycle in Chinese hamster ovary cells. *J. Cell Biol.* **106**, 431–439 (1988).
- Jantsch-Plunger, V. *et al.* CYK-4: A Rho family GTPase activating protein (GAP) required for central spindle formation and cytokinesis. *J. Cell Biol.* **149**, 1391–1404 (2000).
- Powers, J., Bossinger, O., Rose, D., Strome, S. & Saxton, W. A nematode kinesin required for cleavage furrow advancement. *Curr. Biol.* **8**, 1133–1136 (1998).
- Raich, W. B., Moran, A. N., Rothman, J. H. & Hardin, J. Cytokinesis and midzone microtubule organization in *Caenorhabditis elegans* require the kinesin-like protein ZEN-4. *Mol. Biol. Cell* **9**, 2037–2049 (1998).
- Gray, C. H., Good, V. M., Tonks, N. K. & Barford, D. The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase. *EMBO J.* **22**, 3524–3535 (2003).
- Gruneberg, U., Glötzer, M., Gartner, A. & Nigg, E. A. The Cdc14 phosphatase is required for cytokinesis in the *Caenorhabditis elegans* embryo. *J. Cell Biol.* **158**, 901–914 (2002).
- Parry, D. H. & O'Farrell, P. H. The schedule of destruction of three mitotic cyclins can dictate the timing of events during exit from mitosis. *Curr. Biol.* **11**, 671–683 (2001).
- Wheatley, S. P. *et al.* CDK1 inactivation regulates anaphase spindle dynamics and cytokinesis *in vivo*. *J. Cell Biol.* **138**, 385–393 (1997).
- Nislow, C., Lombillo, V. A., Kuriyama, R. & McIntosh, J. R. A plus-end-directed motor enzyme that moves antiparallel microtubules *in vitro* localizes to the interzone of mitotic spindles. *Nature* **359**, 543–547 (1992).
- Berliner, E. *et al.* Microtubule movement by a biotinylated kinesin bound to streptavidin-coated surface. *J. Biol. Chem.* **269**, 8610–8615 (1994).
- Okada, Y. & Hirokawa, N. Mechanism of the single-headed processivity: diffusional anchoring between the K-loop of kinesin and the C terminus of tubulin. *Proc. Natl Acad. Sci. USA* **97**, 640–645 (2000).
- Meijer, L. *et al.* Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* **243**, 527–536 (1997).

- Matulienė, J. & Kuriyama, R. Kinesin-like protein CHO1 is required for the formation of midbody matrix and the completion of cytokinesis in mammalian cells. *Mol. Biol. Cell* **13**, 1832–1845 (2002).
- Minestrini, G., Harley, A. S. & Glover, D. M. Localization of Pavarotti-KLP in living *Drosophila* embryos suggests roles in reorganizing the cortical cytoskeleton during the mitotic cycle. *Mol. Biol. Cell* **14**, 4028–4038 (2003).
- Kikkawa, M. *et al.* Switch-based mechanism of kinesin motors. *Nature* **411**, 439–445 (2001).
- Andrews, P. D. *et al.* Aurora B regulates MCAK at the mitotic centromere. *Dev. Cell* **6**, 253–268 (2004).
- Lan, W. *et al.* Aurora B phosphorylates centromeric MCAK and regulates its localization and microtubule depolymerization activity. *Curr. Biol.* **14**, 273–286 (2004).
- Ohi, R., Sapra, T., Howard, J. & Mitchison, T. J. Differentiation of cytoplasmic and meiotic spindle assembly MCAK functions by Aurora B-dependent phosphorylation. *Mol. Biol. Cell* **15**, 2895–2906 (2004).
- Mollinari, C. *et al.* PRC1 is a microtubule binding and bundling protein essential to maintain the mitotic spindle midzone. *J. Cell Biol.* **157**, 1175–1186 (2002).
- Mello, C. & Fire, A. DNA transformation. *Methods Cell Biol.* **48**, 451–482 (1995).
- Kaitna, S., Mendoza, M., Jantsch-Plunger, V. & Glötzer, M. Incenp and an aurora-like kinase form a complex essential for chromosome segregation and efficient completion of cytokinesis. *Curr. Biol.* **10**, 1172–1181 (2000).
- Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921–929 (1999).
- Patra, D. & Dunphy, W. G. Xe-p9, a *Xenopus* Suc1/Cks protein, is essential for the Cdc2-dependent phosphorylation of the anaphase-promoting complex at mitosis. *Genes Dev.* **12**, 2549–2559 (1998).
- Boyle, W. J., van der Geer, P. & Hunter, T. Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **201**, 110–149 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by a grant from the Austrian Science Foundation and with the support of Boehringer Ingelheim. M.M., V.P. and M.G. would like to thank K. Bartalska for technical assistance, and S. Kaitna and K. Mechtler for help in the initial stages of this project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G. (mglotzer@imp.univie.ac.at).

Structural basis for inhibition of the replication licensing factor Cdt1 by geminin

Changwook Lee¹, BumSoo Hong^{1*}, Jung Min Choi^{1*}, Eugene Kim¹, Saori Watanabe², Yukio Ishimi³, Takemi Enomoto², Shusuke Tada², Youngchang Kim⁴ & Yunje Cho¹

¹National Creative Research Center for Structural Biology and Department of Life Science, Pohang University of Science and Technology, Hyo-ja dong, San31, Pohang, KyungBook, South Korea

²Graduate School of Pharmaceutical Sciences, Tohoku University, Aoba-Ku, Sendai, Miyagi 980-8578, Japan

³Biomolecular and Technology Department, Mitsubishi Kagaku Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511, Japan

⁴Bioscience Division, Structural Biology Center, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

* These authors contributed equally to the work.

To maintain chromosome stability in eukaryotic cells, replication origins must be licensed by loading mini-chromosome maintenance (MCM2–7) complexes once and only once per cell cycle^{1–9}. This licensing control is achieved through the activities of geminin^{10–12} and cyclin-dependent kinases^{9,13,14}. Geminin binds tightly to Cdt1, an essential component of the replication licensing system^{6,15–18}, and prevents the inappropriate reinitiation of replication on an already fired origin. The inhibitory effect of geminin is thought to prevent the interaction between Cdt1 and the MCM helicase^{19,20}. Here we describe the crystal structure of the mouse geminin–Cdt1 complex using tGeminin (residues 79–157, truncated geminin) and tCdt1 (residues 172–368, truncated

Cdt1). The amino-terminal region of a coiled-coil dimer of tGeminin interacts with both N-terminal and carboxy-terminal parts of tCdt1. The primary interface relies on the steric complementarity between the tGeminin dimer and the hydrophobic face of the two short N-terminal helices of tCdt1 and, in particular, Pro 181, Ala 182, Tyr 183, Phe 186 and Leu 189. The crystal structure, in conjunction with our biochemical data, indicates that the N-terminal region of tGeminin might be required to anchor tCdt1, and the C-terminal region of tGeminin prevents access of the MCM complex to tCdt1 through steric hindrance.

The overall structure of the tGeminin–tCdt1 complex resembles an axe, in which tGeminin represents the handle and tCdt1 the blade (Fig. 1a, Supplementary Table 1). tGeminin forms a parallel coiled-coil dimer, with two extended helices (Ha and Hb, residues 95–156) packed together in a left-handed superhelix through van der Waals packing, three hydrogen bonds and four ion-pair interactions (Supplementary Fig. 1). The ten residues immediately N-terminal to the coiled-coil domain form a loop and a short helix. These residues and the N-terminal part of the coiled coil of tGeminin bind to tCdt1. The superhelical twist of the tGeminin dimer provides steric complementarity to the two short N-terminal helices of tCdt1, optimizing the interactions between the two proteins. The C-terminal part of the Hb helix in tGeminin shows a sharp bend from Leu 138 to its C-terminal end, presumably because of crystal packing (Fig. 1a).

tCdt1 forms an α/β structure with six helices and two strands (Fig. 1b, Supplementary Fig. 2). A database search revealed that tCdt1 most closely resembles bacterial replication terminator protein (RTP), which binds to bacterial replicative helicase, DnaB²¹. In particular, tCdt1 resembles the hydrophobic patch surface of RTP, which is crucial to its contrahelicase activity²² (Fig. 1b).

The extensive interface between tGeminin and tCdt1 is predominantly hydrophobic and buries a total of 2,400 Å² (Fig. 2). The interface is best described in two parts from the perspective of tCdt1.

In the first interface, two short helices (H1 and H2) and the loop L1 of tCdt1 make tight contact with the N-terminal part of the coiled-coil dimer of tGeminin in an anti-parallel manner through multiple van der Waals contacts, four hydrogen bonds and two ion pairs, with 1,146 Å² of buried surface area (Fig. 2a). The conserved Pro 181, Ala 182, Tyr 183 and Phe 186 in the H1 helix of Cdt1 have a central role, making van der Waals contact with Ala 106, Glu 109 and Ala 110 of the Ha helix, and Leu 107, Leu 111, Asn 114, Glu 115 and His 118 of the Hb helix of tGeminin. These geminin residues are highly conserved in geminin orthologues (Supplementary Fig. 2a). In addition, Leu 189, Ala 190 and Pro 192 of tCdt1 interact with Ala 106 from the Ha helix of tGeminin, and Arg 104, Leu 107, Tyr 108 and Leu 111 from its Hb helix.

To investigate the importance of the interactions described above, we made several deletion mutants in the N-terminal residues of tCdt1 and measured their binding affinity for tGeminin by isothermal titration calorimetry (Supplementary Table 2). The tGeminin binding affinity of tCdt1^{184–368}, in which the first 12 residues of tCdt1 have been removed, was about 7,000-fold lower than that of tCdt1. Because Pro 181, Ala 182 and Tyr 183 are the only tGeminin-contacting residues deleted in this mutant, this result demonstrates the importance of this conserved triad of tCdt1 in the binding of tGeminin (Fig. 2a). tCdt1^{190–368}, in which the H1 and H2 helices of tCdt1 are removed, does not form a complex with tGeminin, indicating that this first interface might be the primary binding site for tGeminin.

The complex structure also reveals the significance of residues 113–119 of tGeminin in the binding of tCdt1. Multiple point mutations on the middle portion (residues 113–119) of tGeminin reduced the binding affinity for tCdt1 2,000-fold, whereas deletion of the C-terminal region (residues 131–157) of tGeminin did not alter the binding affinity (Supplementary Table 2). Because the aliphatic side chains of these residues of tGeminin make van der

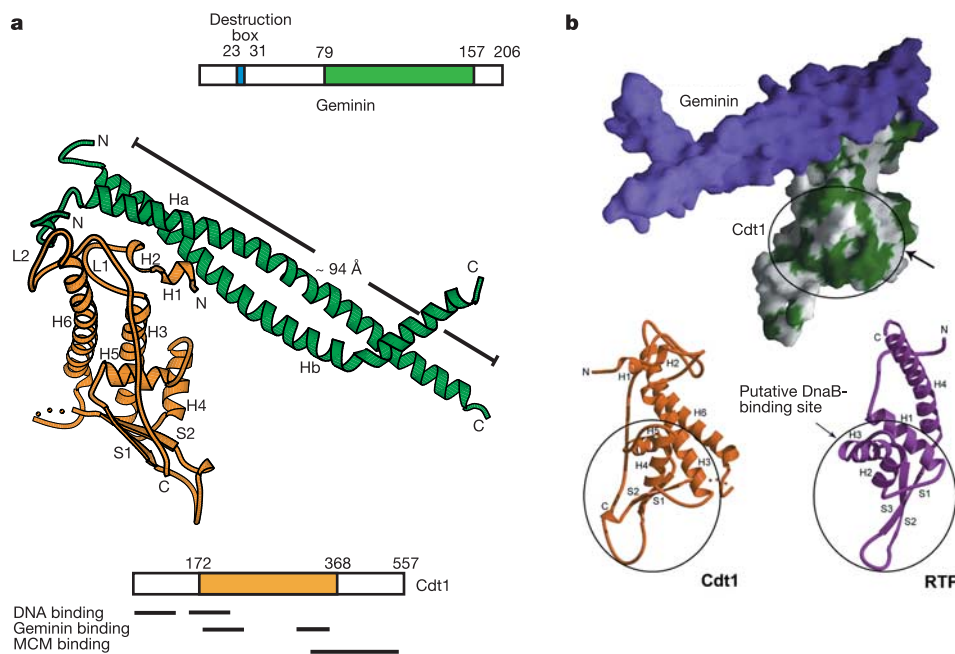


Figure 1 Structure of the tGeminin–tCdt1 complex. **a**, The tGeminin is shown in green and tCdt1 is in orange. The destruction box (blue) and the Cdt1-binding region (green) for geminin are shown above the structure. The boundaries of DNA-binding and MCM2–7-binding regions for Cdt1 are shown below the structure¹⁹. The N-terminal ten residues are structured in the tGeminin monomer that binds to tCdt1 but are disordered in the other tGeminin monomer, indicating that this region might undergo induced folding after binding to tCdt1. **b**, Surface representation of the tGeminin–tCdt1 complex (top). The molecular surface of tCdt1 and tGeminin are coloured white and purple, respectively. The

surface of the tCdt1 residues more than 80% conserved in five Cdt1 orthologues (Supplementary Fig. 2) is coloured in green. The region that resembles the replicative helicase-binding site of RTP^{21,22} is marked with a circle. Comparison of the structures for tCdt1 (left) and RTP (right) is shown at the bottom. The figure is in the same orientation as in the surface representation. The hydrophobic patch surface that mediates the contrahelicase activity of RTP^{21,22} (helices H1 and H2 and strands S1, S2 and S3) is marked with a circle. This region is very similar to the region spanning the helices H3, H4, the intervening loop, and the strands S1 and S2 of tCdt1, which is highly conserved in the Cdt1 family members.

Waals contacts with Pro 181, Ala 182 and Tyr 183 of tCdt1, our data further support the importance of the conserved triad in tGeminin–tCdt1 interaction.

The crucial role of the first interface in the interaction between tGeminin and tCdt1, in which tCdt1 is recognized by both Ha and Hb helices of the tGeminin dimer, indicates that dimeric tGeminin (and by inference dimeric geminin) might be important in preventing replication licensing. In addition, the tGeminin and tCdt1 residues in this interface are highly conserved in a range of metazoan species, indicating that this interface structure might be similar in all metazoan family members (Fig. 2a).

In the second part of the tGeminin–tCdt1 interface, a surface groove on Cdt1 formed by helix H6 and loop L2 of tCdt1 binds the ten residues in the N-terminal loop of tGeminin monomer (Fig. 2b). This region of Cdt1 also partly contributes to its binding to naked DNA¹⁹ (Supplementary Fig. 3). The deletion of these ten residues decreased tGeminin's binding affinity to tCdt1 about 15-fold (Supplementary Table 2). However, this decrease in binding constant is much smaller than those of the first interface mutations, indicating that this interface is likely to have a relatively minor function in the interaction between tGeminin and tCdt1. Consistent with this result was our observation that geminin^{89–157} inhibits the replication in *Xenopus* egg extract, but with about 15-fold less activity than that of tGeminin (Fig. 3a).

How would geminin prevent Cdt1 from loading the MCM2–7 complex on replication origins, thus inhibiting replication? Studies showed that Cdt1 immunoprecipitates together with the MCM complex in yeast²³ and nuclease-treated mammalian cell extracts²⁰ and binds to MCM6 in two-hybrid analysis and to purified MCM4/6/7 complex *in vitro*¹⁹. Geminin and the MCM complex bind to separate sites on Cdt1 (ref. 19). In support of this, the geminin^{79–130} mutant, which binds to tCdt1 with an affinity similar to that of tGeminin, did not interfere with the association between the MCM complex and tCdt1 (Fig. 3b).

Interestingly, residues 140–160 of *Xenopus* geminin (residues 129–149 of mouse geminin) that do not make any contact with tCdt1 in our structure are essential for geminin's ability to inhibit replication¹⁰, suggesting that this part of the extended coiled coil of tGeminin is important for the inhibition of the binding of MCM. To explain the mechanism by which this C-terminal part of geminin

inhibits the loading of MCM, we created several tGeminin mutants (Supplementary Fig. 4) and examined whether these mutant proteins inhibit complex formation between glutathione S-transferase (GST)–Cdt1 and the purified MCM4/6/7 complex, complex formation between GST–Cdt1 and endogenous MCM complex from mouse embryonic fibroblast extracts, and replication in *Xenopus* egg extracts. If the C-terminal part of the extended helix of tGeminin that does not interact with Cdt1 shields the MCM-binding region of Cdt1, changes in the length of the C-terminal portion of the helix should affect the MCM binding activity and the replication inhibitory activity of tGeminin.

When eight residues on the exposed surface of tGeminin were mutated simultaneously, the mutant inhibited DNA replication and MCM binding to GST–Cdt1 in the same manner as wild-type tGeminin (Fig. 3). We next decreased the C-terminal helix length by deleting 4–12 (about 6–17 Å) residues between residues 136 and 151 in tGeminin, and examined the activity of these mutants in the same assays. tGeminin^{Δ136–139} and tGeminin^{Δ144–147} behaved similarly to wild-type tGeminin in both assays. The eight-residue deletion mutant tGeminin^{Δ144–151} also retained both inhibitory activities but tGeminin^{Δ136–147} failed to inhibit DNA replication even when present in a tenfold excess and did not inhibit the binding of MCM to tCdt1, indicating that there might be a threshold length for the C-terminal helix below which it loses its ability to inhibit replication licensing.

Because our data reveal that decreasing the C-terminal helix length in tGeminin affects its ability to inhibit both MCM binding and replication, we further examined the correlation between the activity and the C-terminal length of tGeminin, using a 12-residue (136–147) deletion mutant of geminin^{79–170}. The total number of residues in geminin^{79–170} (Δ136–147) is close to that of wild-type tGeminin or tGeminin^{Δ136–139}. However, this mutant is unlikely to have the continuous straight helix at its C-terminal end seen in tGeminin because this region is easily removed by limited proteolysis (data not shown). Even a tenfold excess of geminin^{79–170} (Δ136–147) did not inhibit DNA replication efficiently (Fig. 3a). However, this mutant is more active than tGeminin^{Δ136–147}, indicating that some activity might have been recovered simply by addition of size.

To confirm that rigidity at the C-terminal end of tGeminin is needed to inhibit the access of the MCM complex, we constructed

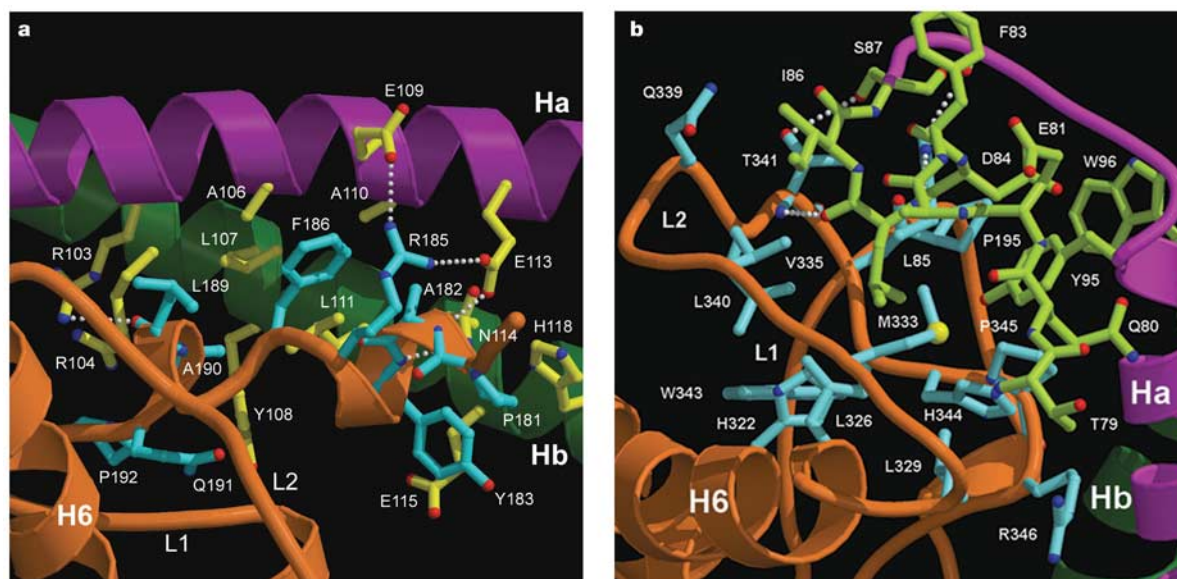


Figure 2 Bipartite tGeminin–tCdt1 interface. **a**, Close-up view of the first interface involving helices H1 and H2 from tCdt1 and Ha and Hb from tGeminin. The secondary structures of tCdt1 (orange), Ha (green) and Hb (pink) of tGeminin, and side chains of tCdt1 (cyan) and tGeminin (yellow) are shown. Oxygen and nitrogen atoms are shown in red and

blue, respectively. The dotted lines indicate intermolecular hydrogen bonds and ion pairs between tCdt1 and tGeminin. **b**, Close-up view of the second interface including the H6 helix and L2 loop from tCdt1, and the N-terminal loop from a tGeminin monomer.

mutants in which N-terminal and C-terminal parts of tGeminin were linked by a flexible linker (6 or 12 residues) or by the helical part of another coiled-coil dimer protein, GCN4 (ref. 24) (Supplementary Information). The flexibly linked mutants were unable to inhibit both replication and MCM binding, but the GCN4-substituted form exhibited the same effects on both assays as tGeminin (Fig. 3). Our mutational analyses suggest that although the C-terminal region of tGeminin acts like a bulky mask to inhibit the binding of MCM to Cdt1, this region must also have a helical structure with a specific orientation to inhibit pre-replicative complex (pre-RC) assembly and DNA replication efficiently.

The extended length and the rigidity of the coiled coil observed in the tGeminin-tCdt1 structure, in conjunction with mutational data, suggest that the C-terminal portion of geminin inhibits the

loading of MCM on Cdt1 through steric hindrance. This model is supported by GST pull-down studies with purified proteins showing that geminin is sufficient to block the binding of the MCM complex to Cdt1 (ref. 19) and by the strong correlation between our *in vitro* binding assay and replication inhibition assay in *Xenopus* egg extract. The strong correlation in our mutational data also suggests that geminin-mediated inhibition of the Cdt1-MCM interaction is important in replication inhibition. However, we cannot conclusively exclude the possibility that geminin might recruit another cellular factor(s) to prevent the MCM complex binding to Cdt1 more efficiently *in vivo*.

The two distinct but tightly coupled roles for tGeminin, direct binding to tCdt1 through its N-terminal part and inhibition of the binding of the MCM complex to Cdt1 through its C-terminal part, begin to explain how geminin inhibits replication licensing, and highlight the structural importance of the extended coiled coil present in the geminin dimer for the inhibition of pre-RC assembly. Because the ability of geminin to inhibit pre-RC assembly can result in cancer-specific cell killing²⁵, the structure presented here might provide a framework for the design of new anti-cancer drugs. □

Methods

Cloning, expression and purification

See Supplementary Information.

Crystallization and structure determination

Crystals of tGeminin-tCdt1 complex were grown at 22 °C by hanging-drop vapour diffusion from 10% poly(ethylene glycol) 6000, bis-tris propane-HCl, 3% t-butanol, 5 mM dithiothreitol, pH 8.5. The crystals formed in space group C2 with $a = 113.9$ Å, $b = 94.5$ Å, $c = 115.5$ Å and contained two complexes (four tGeminin and two tCdt1) in an asymmetric unit. Diffraction data were collected at -170 °C with crystals flash frozen in crystallization buffer containing 30% glycerol.

A single-wavelength data set was collected with a Se-Met crystal on beamline 19ID of the Structural Biology Center at the Advanced Photon Source at Argonne National Laboratory. Integration, scaling and merging of the diffraction data were done with the HKL2000 suite of programs²⁶. Sixteen selenium sites and initial phases were determined with the CNS program²⁷. After density modification, the electron density map calculated to 3.5 Å was of excellent quality. The phases were further improved by twofold averaging, which allowed us to trace most of the chains. Successive rounds of model building with the O program²⁸, simulated annealing refinement with CNS²⁷ and phase combination allowed the complete building of the structure. The final model contains residues 79–157 of tGeminin and residues 179–290 and 295–365 of tCdt1.

Measurement of Cdt1 binding to geminin proteins

The binding constants of Cdt1 proteins to various geminin proteins were measured by ITC with the Micro Calorimetry System (MicroCal Inc.). All samples used in the ITC experiments were dialysed against 25 mM Tris-HCl pH 7.5, 200 mM NaCl, 7 mM 2-mercaptoethanol. The ITC measurements were performed at 18 °C by making 15–30 injections (7 µl each) of the Cdt1 solution into 1.4 ml of geminin proteins. The concentrations of the Cdt1 and geminin proteins were 0.056–0.18 mM and 0.008–0.036 mM, respectively. Curve fitting was performed with MicroCal Origin software.

MCM complex binding assay

Full-length MCM4/6/7 complex was purified as described previously²⁹. Purified GST-Cdt1 (full-length; 50 ng) and various geminin proteins were mixed with the MCM4/6/7 complex and incubated with glutathione-Sepharose for 6 h at 4 °C in 600 µl of PBS. After being washed with PBS, GST-tagged proteins were eluted and then analysed by SDS-polyacrylamide-gel electrophoresis and western blot analysis with anti-MCM6 (Santa Cruz Biotechnology).

Murine embryonic fibroblast (MEF) cells were maintained in DMEM medium supplemented with 10% fetal calf serum. MEF cells (about 10⁶ cells per sample) were harvested by treatment with trypsin, washed once with PBS, and frozen as cell pellets at -80 °C. Cell pellets were thawed in 500 µl of 50 mM HEPES pH 7.5, 0.1 M potassium acetate, 0.02% Nonidet P40, 10% glycerol, 5 mM MgCl₂ supplemented with protease inhibitors, 1 mg ml⁻¹ lysozyme, 1 mM sodium orthovanadate, 25 mM β-glycerol phosphate and 1 mM ATP. CaCl₂ was added to a final concentration of 5 mM, and micrococcal nuclease S7 was added to each cell suspension. Lysates were incubated on ice for 30 min and clarified by centrifugation. Equal amounts of cell extracts were added to each tube and allowed to form complexes with GST-Cdt1 in the presence of various tGeminin mutant proteins at 4 °C. After extensive washing, proteins bound to GST-Cdt1 were detected with anti-MCM6 as described above. A portion corresponding to 10% of the total cell extracts was used for input.

Replication inhibition assay

Xenopus egg extract and chromatin templates were purified as reported previously¹¹. Interphase *Xenopus* egg extract (10 µl) was supplemented with 25 mM phosphocreatine, 15 µg ml⁻¹ creatine phosphokinase, 0.25 mg ml⁻¹ cycloheximide, [α-³²P]dATP (20 kBq)

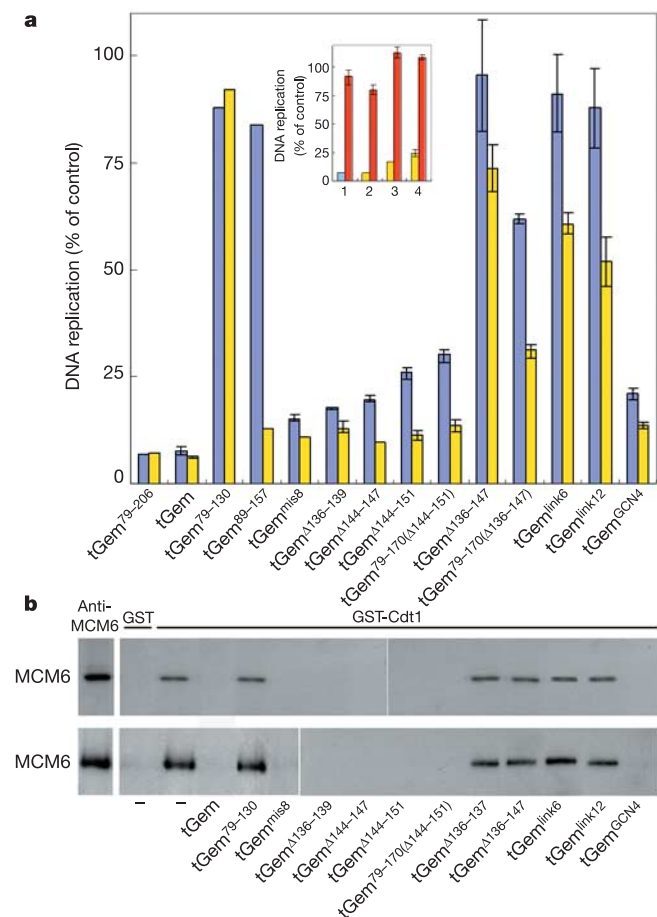


Figure 3 The C-terminal part of tGeminin is important for inhibition of replication and MCM-binding. **a**, Replication inhibition by various geminin mutant proteins. The percentage of DNA replication occurring in the presence of each mutant is indicated. Different concentrations of various geminin mutants were used: blue bars, 1 µM; yellow bars, 10 µM. Eight residues were mutated in tGeminin^{mis8}, and a flexible linker of 6 or 12 residues (KGSREG or KGTREGKGSPEG) was inserted between residues 135 and 136 for tGeminin^{link6} or tGeminin^{link12} (Supplementary Information). For tGeminin^{GCN4}, a 12-residue linker (KDLAEVAEHVQY) from yeast GCN4 (ref. 24) was fused to tGeminin (between residues 135 and 148). The inset shows a rescue experiment showing that the inhibitory effect of the geminin mutants was caused by the specific binding and inhibition targeted to Cdt1. Only results with tGeminin (lanes 1 and 2), tGeminin^{mis8} (lane 3) and tGeminin¹³⁶⁻¹³⁹ (lane 4) are shown here for clarity; other tGeminin mutants exhibited almost identical results. tGeminin proteins at 1 µM (lane 1) or 10 µM (lanes 2–4) were used in the assay and 0.1 µM *Xenopus* Cdt1 (red bar) was added. Results are means ± s.e.m. from triplicate experiments. **b**, The association between Cdt1 and the MCM complex. The purified MCM4/6/7 complex (top) or a nuclease-treated cell extract of MEF cells (bottom) was incubated with GST-Cdt1 (full-length) and various geminin proteins, and detected by western blotting with anti-MCM6.

and the indicated amount of geminin proteins. DNA (3 ng per μ l of demembrated sperm chromatin) was added to the egg extract to start the reaction. After 90 min incubation at 23 °C, the reaction was stopped and the radioactivity incorporated into the acid-insoluble fraction was measured to quantify newly synthesized DNA. For a rescue experiment, 0.1 μ M recombinant *Xenopus* Cdt1 was added to the reaction mixture that was used in replication inhibition assay.

Received 29 March; accepted 5 July 2004; doi:10.1038/nature02813.
Published online 1 August 2004.

1. Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
2. Blow, J. J. & Hodgson, B. Replication licensing—defining the proliferative state? *Trends Cell Biol.* **12**, 72–78 (2002).
3. Diffley, J. F. DNA replication: building the perfect switch. *Curr. Biol.* **11**, R367–R370 (2001).
4. Bell, S. P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128–134 (1992).
5. Coleman, T. R., Carpenter, P. B. & Dunphy, W. G. The *Xenopus* Cdc6 protein is essential for the initiation of a single round of DNA replication in cell-free extracts. *Cell* **87**, 53–63 (1996).
6. Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**, 625–628 (2000).
7. Shreeram, S. & Blow, J. J. The role of the replication licensing system in cell proliferation and cancer. *Prog. Cell Cycle Res.* **5**, 287–293 (2003).
8. Vaziri, C. *et al.* A p53-dependent checkpoint pathway prevents rereplication. *Mol. Cell* **11**, 997–1008 (2003).
9. Kearsey, S. E. & Cotterill, S. Enigmatic variations: divergent modes of regulating eukaryotic DNA replication. *Mol. Cell* **12**, 1067–1075 (2003).
10. McGarry, T. J. & Kirschner, M. W. Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043–1053 (1998).
11. Tada, S., Li, A., Maiorano, D., Mechali, M. & Blow, J. J. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. *Nature Cell Biol.* **3**, 107–113 (2001).
12. Wohlschlegel, J. A. *et al.* Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science* **290**, 2309–2312 (2000).
13. Yamaguchi, R. & Newport, J. A role for Ran-GTP and Crm1 in blocking re-replication. *Cell* **113**, 115–125 (2003).
14. Li, A. & Blow, J. J. Non-proteolytic inactivation of geminin requires CDK-dependent ubiquitination. *Nature Cell Biol.* **6**, 260–267 (2004).
15. Hofmann, J. F. & Beach, D. Cdt1 is an essential target of the Cdc10/Sct1 transcription factor: requirement for DNA replication and inhibition of mitosis. *EMBO J.* **13**, 425–434 (1994).
16. Maiorano, D., Moreau, J. & Mechali, M. XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*. *Nature* **404**, 622–625 (2000).
17. Whittaker, A. J., Royzman, I. & Orr-Weaver, T. L. *Drosophila* double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. *Genes Dev.* **14**, 1765–1776 (2000).
18. Arentson, E. *et al.* Oncogenic potential of the DNA replication licensing protein CDT1. *Oncogene* **21**, 1150–1158 (2002).
19. Yanagi, K., Mizuno, T., You, Z. & Hanaoka, F. Mouse geminin inhibits not only Cdt1–MCM6 interactions but also a novel intrinsic Cdt1 DNA binding activity. *J. Biol. Chem.* **277**, 40871–40880 (2002).
20. Cook, J. G., Chasse, D. A. & Nevins, J. R. The regulated association of Cdt1 with minichromosome maintenance proteins and Cdc6 in mammalian cells. *J. Biol. Chem.* **279**, 9625–9633 (2004).
21. Bussiere, D. E., Bastia, D. & White, S. W. Crystal structure of the replication terminator protein from *B. subtilis* at 2.6 Å. *Cell* **80**, 651–660 (1995).
22. Gautam, A., Mulugu, S., Alexander, K. & Bastia, D. A single domain of the replication termination protein of *Bacillus subtilis* is involved in arresting both DnaB helicase and RNA polymerase. *J. Biol. Chem.* **276**, 23471–23479 (2001).
23. Tanaka, S. & Diffley, J. F. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2–7 during G1 phase. *Nature Cell Biol.* **4**, 198–207 (2002).
24. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539–544 (1991).
25. Shreeram, S., Sparks, A., Lane, D. P. & Blow, J. J. Cell type-specific responses of human cells to inhibition of replication licensing. *Oncogene* **21**, 6624–6632 (2002).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Brünger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Kleywegt, G. J. & Jones, T. A. Efficient rebuilding of protein structures. *Acta Crystallogr. D* **50**, 829–832 (1996).
29. You, Z., Komamura, Y. & Ishimi, Y. Biochemical analysis of the intrinsic Mcm4–Mcm6–Mcm7 DNA helicase activity. *Mol. Cell. Biol.* **12**, 8003–8015 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Son and A. Jeon for help in initial protein purification, Y. Kong for MEF cells, J. Lee for anti-MCM6 antibody, and P. A. Karplus, S. H. Kim, Y. Kong and J. Bradbury for critical readings of the manuscript. This work was supported by the funds from the National Creative Research Initiatives (Ministry of Science and Technology).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.C. (yunje@postech.ac.kr). The coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1WLQ).

Structural basis for redox regulation of Yap1 transcription factor localization

Matthew J. Wood¹, Gisela Storz¹ & Nico Tjandra²

¹Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892-5430, USA

²Laboratory of Biophysical Chemistry, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-8013, USA

The ability of organisms to alter their gene expression patterns in response to environmental changes is essential for viability. A central regulator of the response to oxidative stress in *Saccharomyces cerevisiae* is the Yap1 transcription factor. Upon activation by increased levels of reactive oxygen species, Yap1 rapidly redistributes to the nucleus where it regulates the expression of up to 70 genes^{1–3}. Here we identify a redox-regulated domain of

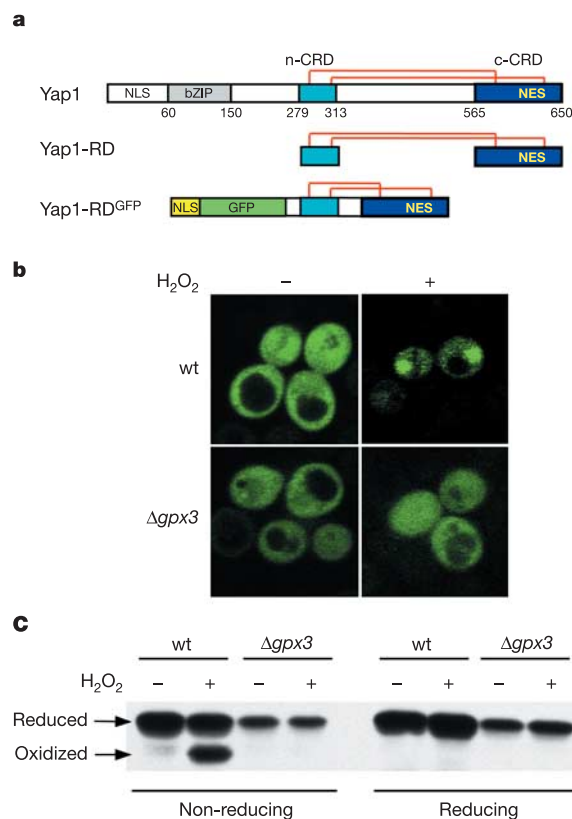


Figure 1 Schematic Yap1 structures and *in vivo* analysis of Yap1-RD^{GFP} subcellular localization and oxidation. **a**, Yap1 contains three conserved regions: a basic leucine zipper DNA binding domain (bZIP), an n-CRD (Asn279 to Arg313) and a c-CRD, (Asn565 to Asn650). The NLS and NES are located at the N and C termini, respectively. The Cys303–Cys598 and Cys310–Cys629 disulfide bonds are shown with red lines. The oxidized Yap1-RD construct used for structure determination consisted of the protease-resistant n-CRD and c-CRD domains. Yap1-RD^{GFP} consisted of an SV40 NLS, GFP and residues Asn279 to Arg313 fused to residues Asn549 to Asn650 of Yap1. This fragment encompasses the n-CRD and c-CRD sequences plus a small amount of the native linker. **b**, Fluorescence microscopy of wild-type and Δ *gpx3* cells expressing Yap1-RD^{GFP} from the native *YAP1* promoter on a *CEN* plasmid. **c**, Oxidized and reduced Yap1-RD^{GFP} extracted from wild-type and Δ *gpx3* cells. Exponentially growing cells were either exposed to H₂O₂ or left untreated. Cell extracts were run on non-reducing and reducing SDS–PAGE gels and probed with a GFP antibody.

Yap1 and determine its high-resolution solution structure. In the active oxidized form, a nuclear export signal (NES) in the carboxy-terminal cysteine-rich domain is masked by disulphide-bond-mediated interactions with a conserved amino-terminal α -helix. Point mutations that weaken the hydrophobic interactions between the N-terminal α -helix and the C-terminal NES-containing domain abolished redox-regulated changes in subcellular localization of Yap1. Upon reduction of the disulphide bonds, Yap1 undergoes a change to an unstructured conformation that exposes the NES and allows redistribution to the cytoplasm. These results reveal the structural basis of redox-dependent Yap1 localization and provide a previously unknown mechanism of transcription factor regulation by reversible intramolecular disulphide bond formation.

Yap1 regulates the transcription of antioxidant defence genes in response to reactive oxygen species such as H_2O_2 . In unstressed cells, Yap1 is freely imported and exported from the nucleus^{4,5}. In cells exposed to H_2O_2 , nuclear export is arrested because Yap1 can no longer interact with the conserved nuclear exporter Crm1 (also known as Xpo1) (ref. 4). This redox control of Yap1 nuclear export requires an N-terminal cysteine-rich domain (n-CRD) and a C-terminal cysteine-rich domain (c-CRD), which encompasses the NES^{1,6} (Fig. 1a).

The oxidized form of Yap1 contains a protease-resistant domain, Yap1-RD, comprised of residues Asn279 to Arg313 of the n-CRD and Asn565 to Asn650 of the c-CRD, covalently attached via Cys303–Cys598 and Cys310–Cys629 disulphide bonds⁷ (Fig. 1a). To determine whether the subcellular localization of this domain is regulated like full-length Yap1 in response to H_2O_2 , we expressed a fusion protein containing a simian virus 40 (SV40) nuclear localization signal (NLS), green fluorescent protein (GFP) and Yap1-RD

(hereafter referred to as Yap1-RD^{GFP}) in *S. cerevisiae* (Fig. 1a). In untreated cells Yap1-RD^{GFP} was localized throughout the cell (Fig. 1b). In cells treated with H_2O_2 , Yap1-RD^{GFP} was relocated to the nucleus within 5 min (Fig. 1b), as is the case with full-length Yap1 (refs 6, 8) (Supplementary Fig. S1). We also examined Yap1-RD^{GFP} localization in a $\Delta gpx3$ null strain. Gpx3 (also known as Orp1) has been shown to regulate Yap1 oxidation *in vivo*⁹. Like the full-length protein, Yap1-RD^{GFP} was no longer redistributed to the nucleus in a $\Delta gpx3$ strain treated with H_2O_2 (Fig. 1b). To confirm that Yap1-RD^{GFP} could be oxidized *in vivo*, we made cell extracts from wild-type and $\Delta gpx3$ strains treated with H_2O_2 or left untreated, and monitored the gel mobility of Yap1-RD^{GFP}. Upon treatment with H_2O_2 , Yap1-RD^{GFP} migrated faster on non-reducing SDS–polyacrylamide gel electrophoresis (PAGE) gels, indicative of an oxidized form of the protein⁸ (Fig. 1c). When these samples were run under reducing conditions the protein reverted back to a slower migrating species, indicative of the reduced form of the protein (Fig. 1c). Extracts prepared from a $\Delta gpx3$ strain showed no Yap1-RD^{GFP} oxidation upon H_2O_2 treatment (Fig. 1c). Thus the redox-dependent regulation of Yap1 could be reconstituted with Yap1-RD.

The ability to recast the redox regulation with Yap1-RD provided an opportunity for mechanistic and structural studies. To elucidate how the subcellular localization of Yap1 is controlled by reduction and oxidation, we determined the high resolution nuclear magnetic resonance (NMR) structure of oxidized Yap1-RD (Fig. 1a). The n-CRD contained a short eight-residue α -helix (n- α 1), whereas the c-CRD contains both β -sheets and α -helices starting at Ser594 and continuing to residue Asn650 (Fig. 2b). Backbone amide ¹⁵N relaxation measurements (T_2) also showed that ~70 amino acids in both the n-CRD and c-CRD are well ordered and that the structured regions correlate with areas of homology and secondary

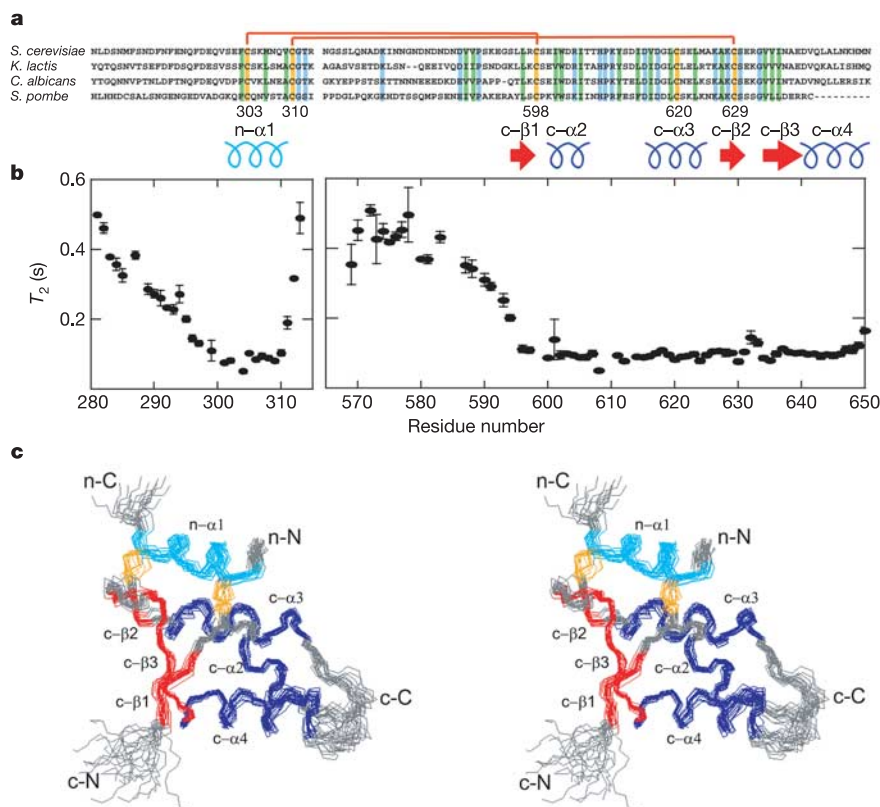


Figure 2 Structural characterization of oxidized Yap1-RD. **a**, Comparison of the n-CRD and c-CRD domains of four Yap1 homologues. The conserved cysteine and hydrophobic residues are highlighted in yellow and green, respectively. All other conserved residues are highlighted in blue. The Cys303–Cys598 and Cys310–Cys629 disulphide bonds are shown with red lines. **b**, The secondary structure and ¹⁵N backbone dynamics (T_2) of

Yap1-RD. Error bars indicate s.d. **c**, Stereo view of the backbone superposition of the ensemble of 20 Yap1-RD NMR structures with only the structured regions shown. n-CRD n- α 1 helix, cyan; c-CRD β -sheets, red; c-CRD α -helices, dark blue; all other regions, grey. The Cys303–Cys598 and Cys310–Cys629 disulphide bonds are shown in yellow. The N and C termini of the n-CRD and c-CRD are indicated with an n- and c-, respectively.

structure (Fig. 2a, b). In addition, the T_2^{average} for the structured regions of the n-CRD and c-CRD are the same, indicating that the two peptides comprising the structured regions of oxidized Yap1-RD behave as a single, well folded protein domain. The 20 lowest energy structures calculated for Yap1-RD have a backbone average r.m.s. deviation of 0.73 Å for the structured regions (Fig. 2c). Both the Cys303–Cys598 and Cys310–Cys629 disulphide bonds are on one side of the Yap1-RD structure, with the side chains of the disulphide-bonded cysteine residues largely solvent-exposed (Figs 2c and 3a).

The residues comprising the NES of Yap1-RD are located on the c- α 3 helix and interact with conserved hydrophobic residues in the n- α 1 helix (Fig. 3a, b). Previous experiments have shown that Leu619 is critically important for Yap1 nuclear export, whereas residues Ile614, Val616 and Leu623 are moderately important^{4,8}. Both Leu619 and Leu623 are buried in the hydrophobic core of Yap1-RD and are completely solvent-inaccessible whereas Ile614 and Val616 are partially exposed (Fig. 3a, b). Importantly, the Val616, Leu619 and Leu623 side chains form extensive hydrophobic contacts with Phe302 and Met306 on the n- α 1 helix (Fig. 3b, c). The

phenyl ring of Phe302 locks Leu619 into the hydrophobic core and also interacts with Val616. The side chain of Met306 directly interacts with Leu623 and inserts into the central hydrophobic cavity of the c-CRD. In addition, Val309 interacts with the c- α 3 helix, but is mostly solvent-exposed and has a modest number of contacts with Met624 and Ala627.

On the basis of the Yap1-RD structure, we proposed that point mutations of residues within the n- α 1 helix would prevent the ability of Yap1 to oxidize and mask its NES. In particular, Phe302 and Met306 seemed to be critical for stabilization of the n-CRD and c-CRD interaction in the oxidized conformation (Fig. 3c). To test our hypothesis that the conserved hydrophobic residues on the n- α 1 helix are critical for Yap1 function, we mutated Phe302, Met306 and Val309 to alanine and analysed Yap1-RD^{GFP} localization and redox state in response to H₂O₂. The F302A and M306A mutants both showed impaired nuclear accumulation and did not oxidize upon H₂O₂ treatment (Fig. 3d, e). In contrast, the V309A mutant showed a phenotype similar to wild-type Yap1-RD^{GFP} (Fig. 3d, e). The F302A and M306A mutants of full-length Yap1 also showed impaired nuclear accumulation and oxidation in response to H₂O₂ treatment, whereas the V309A mutant behaved in a similar way to the wild type (Supplementary Fig. S1). These results show that disulphide bonds are not enough to stabilize the interaction between the n-CRD and the c-CRD, and suggest that masking of the NES requires specific interactions involving both Phe302 and Met306. It is also possible that the Phe302 and Met306 mutations affect the redox potential of the Cys303 and Cys310 residues.

To understand how the reduction of oxidized Yap1 results in unmasking of the NES and redistribution of the protein to the cytoplasm, we treated a ¹⁵N-labelled sample of oxidized Yap1-RD with the reducing agent dithiothreitol (DTT) and monitored changes in its ¹H and ¹⁵N chemical shifts with NMR. Upon reduction, the peaks assigned to structured residues of Yap1-RD completely disappeared and the ¹H–¹⁵N hetero-nuclear single quantum coherence (HSQC) spectrum was characteristic of a protein with little secondary and tertiary structure¹⁰ (Fig. 4a). A comparison of the circular dichroism spectra of oxidized and reduced Yap1-RD also showed a loss of secondary structure upon reduction with DTT (Fig. 4b). Furthermore, size exclusion chromatography indicated that upon reduction, the n-CRD and c-CRD peptides dissociate and migrate separately at molecular weights characteristic of unfolded peptides (data not shown). Biophysical studies using oxidized full-length Yap1 also show conformational changes upon reduction with DTT (ref. 7). Although we cannot rule out the possibility that the c-CRD may adopt a structured conformation in the reduced form as a result of interactions with other

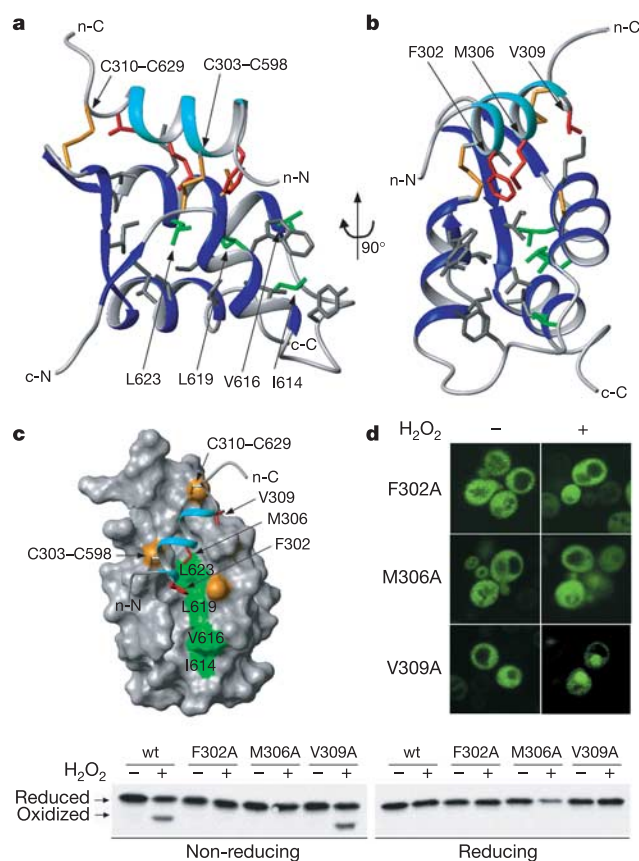


Figure 3 Inhibition of the Yap1 NES by the n- α 1 helix. **a**, Ribbon diagram of the Yap1-RD structure with the lowest energy shown in the same orientation as Fig. 2c. The n- α 1 helix and the regions of c-CRD secondary structure are shown in cyan and dark blue, respectively. The NES residues Ile614, Val616, Leu619 and Leu623 are shown in green. These interact with other hydrophobic core residues of the c-CRD, which are shown in grey. **b**, The same ribbon diagram as in Fig. 3a, rotated to show the n- α 1 residues. The amphipathic n- α 1 helix contains conserved hydrophobic residues, Phe302, Met306 and Val309, shown in red. **c**, Surface representation of the c-CRD domain and its interaction with the hydrophobic residues in the n- α 1 helix. The surface of the NES residues Ile614, Val616, Leu619 and Leu623 are shown in green and Phe302, Met306 and Val309 in red. **d**, Fluorescence microscopy of cells expressing Yap1-RD^{GFP} F302A, M306A and V309A mutants, untreated or treated with H₂O₂ as carried out in Fig. 1b. **e**, Oxidized and reduced Yap1-RD^{GFP} F302A, M306A and V309A mutants extracted from exponentially growing cells untreated or treated with H₂O₂. The cell extracts were prepared as in Fig. 1c.

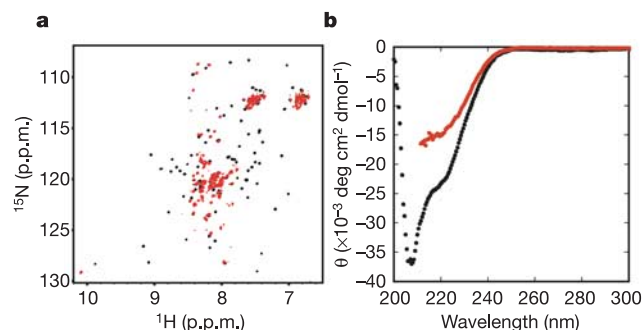


Figure 4 Redox-mediated conformational changes in Yap1-RD. **a**, HSQC spectra of ¹⁵N-labelled oxidized (black) and reduced (red) Yap1-RD. Yap1-RD was reduced by the addition of 20 mM DTT for 10 min before collecting an HSQC spectrum. **b**, Circular dichroism spectra of oxidized (black) and reduced (red) Yap1-RD. All circular dichroism experiments were performed with a 0.1-cm-pathlength cell and 30 μ M Yap1-RD prepared in the same way as the NMR samples.

proteins or portions of Yap1 that are not present in Yap1-RD, our observations suggest that in reduced Yap1 the NES is exposed, allowing Yap1 to be exported from the nucleus by Crm1. Upon oxidation, Yap1 adopts a conformation in which the NES is concealed, as indicated in this structure, allowing Yap1 to accumulate in the nucleus.

To our knowledge, the oxidized Yap1-RD structure is the first known high resolution structure of the sensory domain of a eukaryotic transcription factor that is reversibly regulated by disulphide bond formation. Comparison of Yap1 with other redox-regulated proteins indicates that redox-mediated conformational changes are a general mechanism for regulation of protein function. The formation of a single disulphide bond in OxyR, a transcription factor in *Escherichia coli*, results in restructuring of the protein, leading to a change in DNA binding¹¹. The domain containing the redox active cysteines in the *E. coli* heat shock protein Hsp33 becomes more flexible upon oxidation, allowing association of the Hsp33 dimerization domains and subsequent activation of the chaperone^{12,13}. An intriguing aspect of Yap1 regulation is that the disulphide-bonded cysteine residues are separated by a large ~300-amino-acid flexible domain. The mechanism by which the distant n-CRD and c-CRD domains are rapidly and precisely brought together remains to be determined.

Control of subcellular localization is a well established theme in the regulation of eukaryotic transcription factors^{14,15}. As far as we know, the Yap1 structure is the first example of NES masking mediated by intramolecular disulphide bonds. Other mechanisms by which protein localization signals are masked include phosphorylation of the NLS sequences of NF-AT transcription factors and oligomerization and burial of the NES of the p53 transcription factor^{16–18}. In addition, the dileucine protein-sorting motif of CD4 is masked when it is in a complex with the protein Lck and Zn²⁺, whereas uncomplexed CD4 is unstructured and the dileucine protein sorting motif is exposed¹⁹. We suggest that redox-controlled masking of a signal sequence may represent a general stress-sensitive mechanism for controlling accessibility of protein localization signals. □

Methods

Plasmids, strains and growth conditions

The *S. cerevisiae* parent strain used in this study is YPH499 (*MATa ura3-52 lys2-801^{amber} ade2-101^{ochre} trp1-Δ63 his3-Δ200 leu2-Δ1*). The isogenic Δ gpx3 derivative was made by replacing the coding region of *GPX3* with KanMX (ref. 20). The Yap1-RD-expressing plasmid was made by separate PCR amplifications of the sequences comprising Asn279 to Lys327 (n-CRD) and Gln549 to Asn650 (c-CRD) of Yap1. These PCR reactions were ligated into the pRSET expression vector (Invitrogen). The Yap1-RD expression construct does not code for any non-native amino acids between the n-CRD and c-CRD. The Yap1-RD^{GFP} plasmid was constructed by subcloning PCR fragments comprised of the *YAP1* promoter and SV40 NLS, enhanced GFP, Yap1-RD and *CYC1* terminator into the pRS316 yeast *CEN* vector (primer sequences available on request)²⁰. Yap1-RD^{GFP} n-CRD point mutants were made using standard oligonucleotide PCR-based mutagenesis procedures. All yeast strains were grown at 30 °C in minimal media containing 0.67% (w/v) yeast nitrogen bases, 2% (w/v) glucose and amino-acid dropout mix supplemented with adenine.

Fluorescence microscopy

Exponentially growing cells carrying Yap1-RD^{GFP} constructs were treated with H₂O₂ for 10 min and analysed with a confocal microscope system (model LSM 510; Carl Zeiss MicroImaging, Inc.) using the 488-nm line.

Preparation of cell extracts

Cell extracts were prepared from exponentially growing cells carrying the Yap1-RD^{GFP} constructs. The cells were treated with H₂O₂ for 5 min and prepared as previously described⁸. Yap1-RD^{GFP} samples were run on 8% SDS-PAGE gels, transferred to nitrocellulose and probed with anti-GFP monoclonal antibodies (Roche Applied Science).

Preparation of Yap1-RD

Yap1-RD was expressed in *E. coli*, purified on Ni-NTA (Qiagen) and Mono-Q (Amersham Pharmacia) columns and oxidized as previously described⁷. Oxidized Yap1-RD was digested with a limiting amount of trypsin for 5 h and purified using reverse phase high-performance liquid chromatography. After digestion and purification, the masses of the n-CRD and c-CRD peptides comprising Yap1-RD showed that they consisted of Yap1

residues Asn279 to Arg313 and Asn565 to Asn650, respectively. The Cys303–Cys598 and Cys310–Cys629 disulphide-bonding pattern of Yap1-RD was confirmed using matrix-assisted laser desorption/ionization mass spectrometry. Isotopically labelled proteins were prepared from cells grown in minimal media containing ¹⁵NH₄Cl and unlabelled or ¹³C-labelled glucose²¹.

NMR spectroscopy

NMR data were acquired on Bruker 600- and 800-MHz spectrometers equipped with triple resonance probes or a cryoprobe. All experiments were performed with ~0.8 mM Yap1-RD at 303 K in 10 mM sodium phosphate (pH 6.0), containing 20 mM NaCl and 10% ²H₂O. All spectra were processed with NMRPipe and analysed with PIPP (refs 22, 23). We obtained backbone resonance assignment using standard triple-resonance experiments²⁴. Four-dimensional ¹⁵N/¹³C-edited and ¹³C/¹³C-edited nuclear Overhauser enhancement (NOE) experiments were used to obtain NOE assignments and distance restraints²⁴. The presence of disulphide bonds was confirmed by the observation of numerous contacts between Cys303 and Cys598 as well as Cys310 and Cys629. No contacts to other cysteine residues were observed for Cys620. The ¹³C β chemical shifts for Cys303, Cys310, Cys598 and Cys629 were consistent with oxidized cysteines and the ¹³C β chemical shift of Cys620 was consistent with a reduced cysteine²⁵. The TALOS program was used to obtain phi and psi backbone dihedral restraints²⁶. NMR experiments to measure residual dipolar couplings were performed on phage-containing samples²⁷.

Structure calculations

Peak intensities from nuclear Overhauser enhancement spectroscopy experiments were translated into a continuous distribution of interproton distances. Structures of Yap1-RD were calculated by a distance geometry and simulated annealing protocol with the incorporation of ¹⁵N–¹H and ¹³C α –¹H dipolar coupling restraints using XPLOR-NIH (refs 28, 29). Structural statistics for the ensemble of 20 Yap1-RD structures are listed in Supplementary Table 1.

Received 24 March; accepted 28 June 2004; doi:10.1038/nature02790.

- Kuge, S., Jones, N. & Nomoto, A. Regulation of γ AP-1 nuclear localization in response to oxidative stress. *EMBO J.* **16**, 1710–1720 (1997).
- Godon, C. *et al.* The H₂O₂ stimulon in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **273**, 22480–22489 (1998).
- Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
- Yan, C., Lee, L. H. & Davis, L. I. Crm1p mediates regulated nuclear export of a yeast AP-1-like transcription factor. *EMBO J.* **17**, 7416–7429 (1998).
- Isoyama, T., Murayama, A., Nomoto, A. & Kuge, S. Nuclear import of the yeast AP-1-like transcription factor Yap1p is mediated by transport receptor Pse1p, and this import step is not affected by oxidative stress. *J. Biol. Chem.* **276**, 21863–21869 (2001).
- Coleman, S. T., Epping, E. A., Steggerda, S. M. & Moye-Rowley, W. S. Yap1 activates gene transcription in an oxidant-specific fashion. *Mol. Cell. Biol.* **19**, 8302–8313 (1999).
- Wood, M. J., Andrade, E. C. & Storz, G. The redox domain of the Yap1p transcription factor contains two disulfide bonds. *Biochemistry* **42**, 11982–11991 (2003).
- Delaunay, A., Isnard, A. D. & Toledano, M. B. H₂O₂ sensing through oxidation of the Yap1 transcription factor. *EMBO J.* **19**, 5157–5166 (2000).
- Delaunay, A., Pflieger, D., Barrault, M. B., Vinh, J. & Toledano, M. B. A thiol peroxidase is an H₂O₂ receptor and redox-transducer in gene activation. *Cell* **111**, 471–481 (2002).
- Demarest, S. J. *et al.* Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553 (2002).
- Choi, H. *et al.* Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105**, 103–113 (2001).
- Graumann, J. *et al.* Activation of the redox-regulated molecular chaperone Hsp33—a two-step mechanism. *Structure (Camb)* **9**, 377–387 (2001).
- Vijayalakshmi, J., Mukherjee, M. K., Graumann, J., Jakob, U. & Saper, M. A. The 2.2 Å crystal structure of Hsp33: a heat shock protein with redox-regulated chaperone activity. *Structure (Camb)* **9**, 367–375 (2001).
- Kaffman, A. & O'Shea, E. K. Regulation of nuclear localization: a key to a door. *Annu. Rev. Cell Dev. Biol.* **15**, 291–339 (1999).
- Kau, T. R., Way, J. C. & Silver, P. A. Nuclear transport and cancer: from mechanism to intervention. *Nature Rev. Cancer* **4**, 106–117 (2004).
- Zhu, J. *et al.* Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEK1. *Cell* **93**, 851–861 (1998).
- Clore, G. M. *et al.* High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* **265**, 386–391 (1994).
- Stommel, J. M. *et al.* A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. *EMBO J.* **18**, 1660–1672 (1999).
- Kim, P. W., Sun, Z. Y., Blacklow, S. C., Wagner, G. & Eck, M. J. A zinc clasp structure tethers Lck to T cell coreceptors CD4 and CD8. *Science* **301**, 1725–1728 (2003).
- Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
- Suzuki, M., Youle, R. J. & Tjandra, N. Structure of Bax: coregulation of dimer formation and intracellular localization. *Cell* **103**, 645–654 (2000).
- Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
- Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. A common-sense approach to peak picking in 2-dimensional, 3-dimensional, and 4-dimensional spectra using automatic computer-analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220 (1991).
- Clore, G. M. & Gronenborn, A. M. Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* **239**, 349–363 (1994).
- Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego, 1995).

26. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
27. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065–1074 (1998).
28. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar ^1H – ^{15}N and ^1H – ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732–738 (1997).
29. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank E. Andrade for preparing the Yap1-RD expression construct, S. Moye-Rowley for providing reagents, C. Jackson for suggestions, C. Wu for use of the mass spectrometer and C. A. Combs for his expertise and advice regarding microscopy-related experiments. We also thank A. Gronenborn, R. Hegde, E. Komives, E. Korn, S. Moye-Rowley and W. Outten for critical reading of this manuscript. M.J.W. is supported by a Research Associateship from the National Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.S. (storzg@mail.nih.gov) or N.T. (nico@helix.nih.gov). Yap1-RD coordinates have been deposited with the Protein Data Bank under accession number 1SSE.

erratum

No stellar p-mode oscillations in space-based photometry of Procyon

Jaymie M. Matthews, Rainer Kuschnig, David B. Guenther, Gordon A. H. Walker, Anthony F.J. Moffat, Slavek M. Rucinski, Dimitar Sasselov & Werner W. Weiss

Nature **430**, 51–53 (2004).

In this Letter, Rainer Kuschnig's surname was misspelled as 'Kusching' in the author list. In addition, the numbering of the

reference list was incorrect. References 1 to 26 should be, respectively: 1, 10–16, 2, 3, 17–26 and 4–9. In ref. 23 (ref. 6), 'Rettler' should read 'Retter'. □

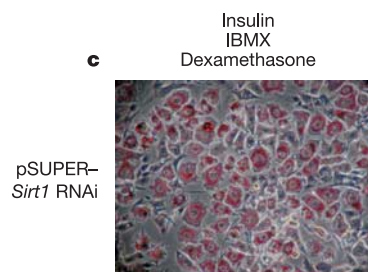
corrigendum

Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR- γ

Frédéric Picard, Martin Kurtev, Namjin Chung, Acharawan Topark-Ngarm, Thanaset Senawong, Rita Machado de Oliveira, Mark Leid, Michael W. McBurney & Leonard Guarente

Nature **429**, 771–776 (2004).

It has been drawn to our attention by Vincent Keng that the image in the bottom-left frame of Fig. 1c of this Letter presents identical data to the one above it on the right. A mistake made by the authors during compilation of Fig. 1 caused the wrong bottom-left image to be used instead of the correct image, which is shown below. The results presented in this replacement micrograph do not alter the conclusions of our study. □



malaria



Cover

A relative takes two malaria-stricken cousins for treatment in Chad.
D. Telemans/Panos

Editor, Nature:
Philip Campbell

Supplement Editors:
Declan Butler
Julie Clayton
Nicola Jones

Supplement Publisher:
Sarah Greaves

Subeditors:
James Kingsland
Colin Sullivan

Production Editor:
Alison Hopkins

Art Editor:
Martin Harrison

Diagrams/Layout:
Alisdair Macdonald

Picture Researcher:
Barbara Izdebska

Production:
Susan Gray

Sponsorship:
Claire Hines

Sponsorship Layouts:
Suzanne Coleman

Marketing:
Claire Aspinall

The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel 44 (0) 20 7833 4000
e-mail: nature@nature.com
http://www.nature.com/
nature



nature publishing group

© 2004 Nature Publishing Group

“Unless we fully embrace country and community ownership, we will be inviting chaos ten years down the road. We need to accelerate

investments in desperately needed vaccines, as well as better treatments. Now is the time to combine long-term investments with crisis management.”

So said Peter Piot, executive director of UNAIDS, at the close of last month's International AIDS Conference in Bangkok. Piot was talking about HIV and AIDS, but his words apply equally well to malaria.

Both diseases kill millions — mostly in Africa, the focus of this supplement. But, compared with AIDS, malaria is neglected by researchers, drug companies, activists and the media. Piot's charismatic leadership is also sorely lacking in the fight against malaria, which is rife with sluggish and ineffective bureaucracy. The drive to fight AIDS has momentum and direction; malaria's seems to be heading for a dead end.

In the following pages, experts reveal that there is greater hope of beating malaria now than ever before. But they also express deep dissatisfaction with the pace and effectiveness of current projects. Patchy international efforts need to be scaled up and given fresh impetus and direction, they say. And billions, not millions, of dollars are needed — one bold move would be to divert some of developing countries' debt repayments to malaria instead.

African countries must also take the fight against malaria into their own hands. Too many big, external players, such as the United Nations' Roll Back Malaria initiative, place too much emphasis on implementing malaria control measures, in the process neglecting strategies to improve local research and infrastructure. Comparable global efforts here would fill a large gap.

Woody Allen famously said that “80% of success is showing up”. Scientists, statesmen and specialists: it's time to show up and demand results.

We are pleased to acknowledge financial support from a number of sponsors: MMV, EDCTP, Impact Malaria, GSK, MIM, MVI, NIAID and Novartis. As always, *Nature* carries sole responsibility for all editorial content. This supplement is also associated with the worldwide launch later this autumn of *Fever Road*, a documentary by London-based Films of Record Ltd.

Philip Campbell and Declan Butler

A collection of *Nature's* material on malaria can be found on our website at www.nature.com/nature/focus/malaria



925 Plague of my people
Pascoal Mocumbi

926 Between hope and a hard place
Brian Greenwood

928 Power to the people
Declan Butler

930 An attack on all fronts
Richard Klausner and Pedro Alonso

932 Where did it all go wrong?
Amir Attaran

934 The invisible victims
Robert Snow

935 Struggling to make an impact
Apoorva Mandavilli

936 Taking aim at mosquitoes
Janet Hemingway

937 The long and winding road
Julie Clayton and Declan Butler

938 Strength in unity
Carter Diggs, Sarah Ewart
and Melinda Moree

940 Save the children
Stephen Hoffman

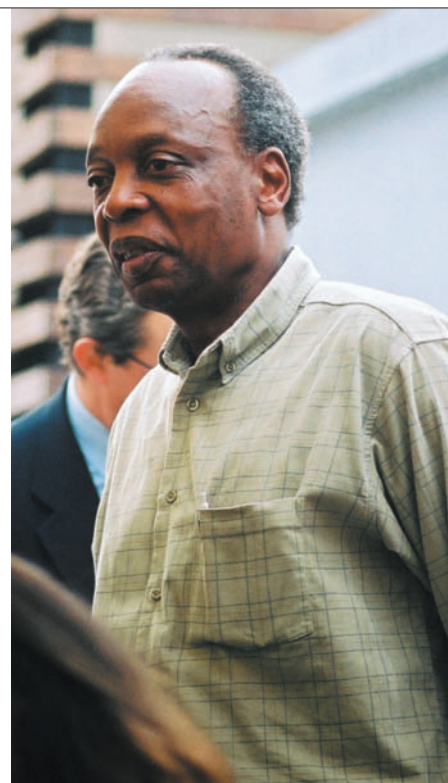
942 Winning the drugs war
Robert Ridley and Yeya Toure

944 Know thine enemy
Daniel Carucci



“It is high time we addressed the widening inequities that characterize our planet today. We need to focus our energies towards achieving basic healthcare for all.”

Pascoal Mocumbi, former prime minister of Mozambique



Plague of my people

In my country, Mozambique, the entire population is at risk of malaria. One in every four children dies before the age of five and, as in much of Africa, malaria is the biggest cause. The poorest populations are most at risk and the disease often strikes children, women who lose their acquired immunity to malaria during pregnancy, and people who lack any immunity to the disease, such as displaced persons and migrants.

Worldwide, malaria is among humanity's largest — and oldest — health and developmental challenges. It kills more than a million people every year and another 300 million to 500 million people have less severe forms of the disease. Half of the world's population lives in the 103 countries where the whine of a mosquito can herald sickness and death.

Malaria also traps countries in a vicious cycle of poverty and ill health. Sub-Saharan Africa loses an estimated US\$12 billion every year from its already meagre gross domestic product because of the disease. Progress by African countries in expanding their economies and reducing poverty over the past decade is now endangered by a failure to use our collective knowledge and wisdom to bring medical advances to bear for the benefit of all our citizens.

Many people in Africa consider malaria and its associated suffering to be an inevitable part of everyday life. Parents live in constant fear of this 'fever'. But malaria is an eminently preventable and curable disease if only adequate measures are taken, diagnosis is prompt and effective treatments are available. One day, vaccines

might provide further help to defeat this deadly disease.

Malaria was only recently labelled a tropical disease, after it was successfully controlled in temperate zones in the 1950s and 1960s. Eradication remains a distant dream in Africa. Today sub-Saharan Africa shoulders 90% of the worldwide malaria burden. The international community cannot with a clear conscience afford to ignore this public-health crisis any longer — in particular now that there is compelling evidence that public health is intimately linked to development and escaping poverty.

Malaria is returning to regions from which it had disappeared, and drug-resistant parasites are emerging — as are insecticide-resistant mosquitoes. These signs should be a wake-up call to politicians and policy-makers, and indeed to the public, as to the urgency of working together to develop research and control strategies to contain this scourge.

Radical approach

Above all, we need to think in radical new ways — ways that show we are more conscious of our common humanity in this third millennium. Ways that make the obligation to help our weakest members the rule and the priority, rather than the exception. And ways in which there is no longer any place for the 'my problem versus your problem' attitude that is still too prevalent in today's otherwise globalized world.

The benefits of scientific innovation must be more equitably shared. Our world is an incredibly complex system, but information and telecommunications technology

has placed us all in a 'small common hut'. Crucial market-driven, economic transformations must be managed so that they also respect fundamental moral values. It is high time we addressed the widening inequities that characterize our planet today. We need to focus our energies towards achieving basic healthcare for all.

If shared more equitably, humankind's knowledge and resources could enable individuals, communities and nations to take responsibility for their own health and developmental challenges. Recently, new international health initiatives such as the Medicines for Malaria Venture, and the European and Developing Countries Clinical Trials Partnership have begun to accelerate the discovery, development and delivery of new tools to control malaria and other poverty-related diseases.

These are a good start. But such initiatives are often themselves grossly underfunded, while funding and other resources are an order of magnitude short of what is needed for a global push that would make a big impact on the disease.

This *Nature* supplement should be commended. It highlights the impact of malaria, the efforts of scientists to defeat it, and the political and economic issues that surround it. I hope these articles will inspire greater awareness and more sharply focused, inclusive global health initiatives to address the growing problem of malaria and other neglected diseases. ■

Pascoal Mocumbi was prime minister of Mozambique from 1994 to February 2004, and is High Representative of the European and Developing Countries Clinical Trials Partnership Programme.

Between hope and a hard place

Campaigns against malaria are multiplying, but so are malaria deaths. Brian Greenwood asks what can be done to turn the tide.

After decades of neglect, the international community is showing renewed interest in malaria. Rich countries are acknowledging that rampant, drug-resistant malaria threatens not only their own citizens when they travel to malaria-endemic areas, but also damages the economies of tropical countries and potential markets for their goods and services¹.

This is a good time to take stock of the situation. How bad is the problem, what is being done to combat the disease, and how is fundraising progressing? There are promising signs on all these fronts, but major obstacles remain. A distinct lack of accurate data on disease burden is hampering the scientific assessment of the impact of control measures. At the same time, a shortage of expertise and infrastructure in the healthcare and research systems of poor countries is threatening their ability to implement international initiatives.

No one knows precisely what the clinical or economic burden is. The official line has long been that 'malaria kills more than 1 million people a year', mostly children, and as a slogan this has been a powerful rallying call in moving malaria up the international agenda.

Whether it is half-a-million child deaths annually or two million, the toll is unacceptable. But to capitalize on the current drive to improve control, getting better statistics on every aspect of malaria is crucial. Estimates of numbers and distribution of malaria deaths have progressed from 'back-of-the-envelope' calculations to sophisticated modelling and geographical information systems that map populations and their levels of risk².

That said, most countries where malaria

is endemic (see map, above) lack a national births and deaths register. In addition, many deaths occur in people's homes without any clinical investigation or diagnosis, making it impossible for demographic surveys to provide precise mortality figures. This in turn hampers accurate assessment of malaria-control programmes (see 'The invisible victims', page 934) — although in many African countries malaria is such an important cause of childhood deaths that overall mortality rates can be used to measure the effectiveness of control measures such as insecticide-treated bednets³. Nonetheless, using various data sources we can make rough estimates of the overall malaria burden by country in Africa⁴.

Front-line attack

Why are so many people still dying from malaria when effective tools are available to control it? One reason is that, despite much noise, the international community and malaria-endemic countries have yet to implement existing tools widely enough (see 'An attack on all fronts', page 930). Experience in Vietnam and a few African countries, such as Eritrea, shows what can be achieved — these nations are running strong and effective national programmes.

In most countries affected by malaria, the lynch-pin of control is readily accessible: effective treatment to prevent deaths. Unfortunately, in much of Africa treatment is anything but accessible and effective. Clinics are few and remote, and resistance has rendered first-line drugs such as chloroquine and sulphadoxine-pyrimethamine ineffect-

ive (see 'Winning the drugs war', page 942).

'Home treatment' options are being explored to provide quicker access to therapies. These include training mothers and community volunteers to treat malaria, improving the therapeutic skills of the local shopkeepers from whom most people in Africa obtain their antimalarial drugs, and evaluating suppositories for the front-line treatment of severe malaria in places where there are no staff trained to give injections.

Artemisinin-based combination therapies (ACTs), derived from the herb *Artemisia annua*, provide a rapid cure and are an immediate solution to the problem of drug resistance. But ACTs cost several times as much as existing drugs, so some aid donors have been reluctant to support their widespread introduction.

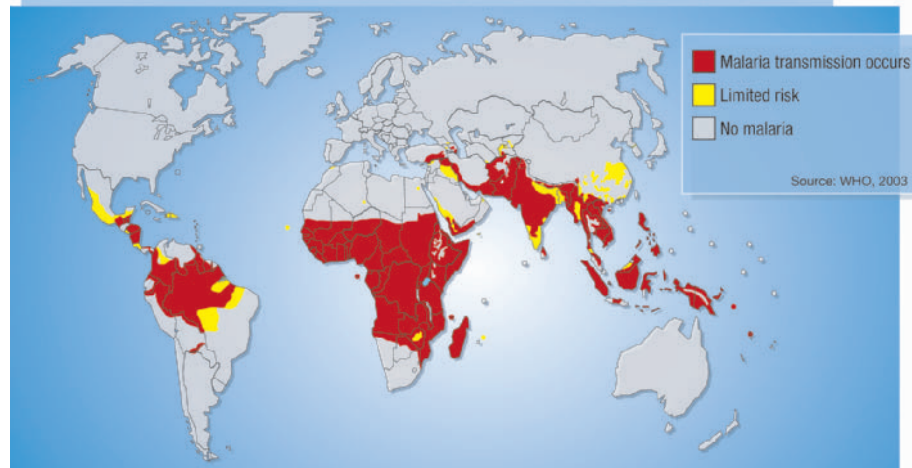
The recent decision by the Global Fund to Fight AIDS, Tuberculosis and Malaria to switch its malaria grants for all African countries to support the use of artemisinins is good news. But relying on treatment as the mainstay of control is an admission of defeat: the

failure to prevent infection in the first place. Better prevention strategies must be the goal.

Spraying the interior of houses with insecticides that leave a lasting residue, insecticide-treated bednets and

intermittent treatment in pregnancy with sulphadoxine-pyrimethamine can all reduce the incidence of new infections. But such measures are not being used widely enough. Few children sleep under bednets⁵, because of their costs, difficulties in getting them to remote areas and ignorance about their benefits. We need to find ways to increase the

At risk: worldwide incidence of malaria





G. OSODI/AP

use of bednets, for example by distributing them through antenatal and infant immunization clinics.

Existing control measures must be supplemented or replaced with new tools. It takes more than a decade to develop new drugs, so we should be priming the pipeline now with molecules that have new mechanisms of action and are less susceptible to resistance. Here, prospects are much more promising than five years ago, thanks to organizations such as the Medicines for Malaria Venture, a Geneva-based body set up to discover and develop new drugs.

Work in progress

New insecticides are also on the horizon. Current bednets use one of the pyrethroid group of insecticides, but the main mosquito vectors *Anopheles gambiae* and *A. funestus* are already developing resistance to these (see 'Taking aim at mosquitoes', page 936).

Substantial progress is being made in the development of malaria vaccines, and in the creation of genetically modified mosquitoes that are resistant to malaria⁵. But both are still far from being ready for widespread use (see 'Save the children', page 940).

In parallel, work on the parasite and vector genomes will ultimately provide a better

understanding of the biology of the *Plasmodium* parasite and aid the discovery of weaknesses in its armour. This will spur the rational design of a new generation of drugs and vaccines. But it is going to be a long, hard battle before we see the fruits of genomics and proteomics turned into products that save lives (see 'Know thine enemy', page 944).

Despite a lack of progress on some fronts, prospects for rolling back malaria look more encouraging in 2004 than at any time since the global malaria eradication campaigns of the 1950s and 1960s. Poor countries have access to increasingly large sums of money for control from international organizations such as the Global Fund. Malaria research is now better funded than ever, thanks to new donors such as the Bill & Melinda Gates Foundation, which has injected several hundred million dollars into the field. Some established donors have doubled or even trebled their funding.

Meanwhile, an essential component of control risks being forgotten: the human resources needed for researching and implementing control measures. The benefits of increased funding will not be fully realized if control efforts continue to rely on poorly trained and paid staff (see 'Power to the people', page 928).

Trained staff are in woefully short supply in all malaria-endemic countries, at every level, from community workers and public-health specialists to researchers. And those health personnel that do exist are being asked to do more and more as additional funds become available for the

Bitten by the bug: posters promote insecticide-treated bednets, but cost and poor distribution mean nets are rarely used in remote areas.

control and treatment of other major infections such as HIV. Many aspects of research can only be tackled in endemic areas, but there is a drastic shortage of support for either research institutions or staff to achieve this.

Organizations such as the World Health Organization's Roll Back Malaria initiative, the Gates Malaria Partnership and the Multilateral Initiative on Malaria are attempting to address the situation, but their efforts are fragmented and insufficient. A major push is needed to tackle the problem coherently across Africa and elsewhere, but it is not clear who will be up to the job (see 'Where did it all go wrong?', page 932). The World Health Organization lacks the funds, and some major donors are nervous about investing in what may seem like a bottomless pit. But if this challenge is not taken up, new tools and the increasing sums of money that are being made available risk being wasted and put into the field with insufficient thought.

A new international impetus is now needed to prevent this happening. Otherwise, the present period of hope for malaria research and control may turn out to have been an illusion.

Brian Greenwood is director of the Malaria Centre at the London School of Hygiene and Tropical Medicine, UK.

1. Sachs, J. & Malaney, P. *Nature* **415**, 680–685 (2002).
2. Korenromp, E. L., Williams, B. G., Gouws, E., Dye, C. & Snow, R. W. *Lancet Infect. Dis.* **3**, 349–358 (2003).
3. Phillips-Howard, P. A. et al. *Am. J. Trop. Med. Hyg.* **68** (Suppl.), 23–29 (2003).
4. *Africa Malaria Report 2003* (WHO/UNICEF, Geneva, 2003); available at www.rbm.who.int/amd2003/amr2003/amr_toc.htm
5. Coleman, P. G. & Alpey, L. *Trop. Med. Int. Health* **9**, 433–437 (2004).

Malaria is estimated to cause each year:

- More than 1 million deaths
- Up to 500 million attacks of acute illness
- Up to 50,000 cases of neurological damage
- Up to 400,000 episodes of severe anaemia in pregnancy
- Up to 300,000 low-birthweight babies





Power to the people

In Africa, where malaria hits hardest, scientists are crying out for countries to take matters into their own hands, says Declan Butler.

Grace Malenga's frustration comes through loud and clear: "Come the first rains, after a week or so, my wards are flooded by sick, convulsing or anaemic kids — a good proportion of whom die, year in, year out. What preventative measures will it take to reverse that?"

As a paediatrician and director of the Malaria Alert Centre in Blantyre, Malawi, Malenga finds that decades of medical progress and international effort have changed little for those in the front line of the war against malaria.

Blantyre is surrounded by tea fields, forested slopes and some of the most striking mountains in Malawi. But during the rains from November to March, malaria turns this region into killing fields. Its sad claim to fame is the international Blantyre Coma Scale, invented here in 1987, which rates the clinical severity of malaria in children.

Malawi, a tiny country at the southern

Ground work: an aid worker explains the malaria life cycle to villagers.

tip of the African Rift Valley, has conquered many other diseases. Its coverage of common childhood vaccines now stands at 80%. Yet malaria is getting worse. Speak to any scientist in Africa and you get the same message: to turn the tide, African doctors and researchers must take matters into their own hands, translating research and control measures into sustainable local campaigns. More needs to be done on the ground, they say. And for this to happen, they need more political will and money, and more local talent.

Drugs and vaccines must also be developed and tested in Africa, experts say. They can't just be shipped in. Their effectiveness can only be assessed on large trial populations in areas where the disease is endemic, and under realistic conditions of use.

"International efforts are positive, but they can drive research in a direction that does not address the nitty-gritty of effective use of research findings for human benefit," says Wilfred Mbacham, a molecular biologist at the University of Yaoundé in Cameroon.

Sweeping decisions by big international programmes can be misguided, says Richard Tren, director of the South African pressure group Africa Fighting Malaria. The move by agencies such as the World Health Organiza-

tion to focus on insecticide-treated bednets while neglecting insecticide spraying, for example, is widely considered a mistake. Spraying, despite its bad environmental image, has been vital to effective control in many countries, says Tren.

People on the ground are best placed to decide which tools are the most effective in areas plagued by mosquitoes and disease, he argues. Take the Konkola Copper Mine, which in 2000 began a campaign to protect its workers and the community in the Konkola highland copper belt in northern Zambia against infection.

Mine of information

To help design the programme, the mine called in Brian Sharp, head of malaria research at the Medical Research Council in South Africa, who runs a successful control programme covering 20,000 km² in Swaziland, Mozambique and South Africa. On Sharp's advice, the mine launched a massive publicity campaign to win public acceptance of spraying and the taking of blood samples. It then taught a team of local workers to spray all the 31,500 houses in a 2,700-km² zone. In one year, they halved the number of malaria cases among some 350,000 people.

L. BERRY/MAGNUM PHOTOS



Brian Sharp (left) believes that gaining the support of local people is key to winning wider acceptance of control measures such as spraying and bednets (right).

The South African-born Sharp says that winning the wholehearted enthusiasm of local people is key to successful malaria control.

In contrast, complains one African scientist, too many international efforts involve foreign consultants who “fly in and out, don’t understand our problems, dictate and control health policy and strategies, and then blame us when they fail”.

The biggest local impact can be made by promoting African research, says Nigerian Solomon Nwaka, a scientific officer of the Medicines for Malaria Venture in Geneva. Progress is being made in training a new generation of African researchers. At an international malaria conference in Tanzania two years ago, half of the thousand delegates were young African PhD students.

Local support

The impetus has come largely from the Multilateral Initiative on Malaria (MIM) — an alliance of research agencies, charities, aid donors and scientists set up in 1997. Its grants are the first substantial international effort to provide support specifically for young African scientists.

Francine Ntoumi, a researcher at Schweitzer Hospital in Lambaréné, Gabon, recalls that as recently as 1998 her laboratory had few staff and no international collaborations. Thanks to MIM grants for research on the genetics of malaria, she now has a team of eight scientists and collaborations with four northern research institutes.

But such grants remain few and small. MIM grants, for example, average just US\$2 million annually. “We have come some of the way, but we are far from achieving the critical mass we require to meet needs in endemic countries,” says Fred Binka, an epidemiologist at the University of Ghana. Schemes to train staff for the day-to-day running of malaria-control programmes are also lacking, says Mbacham.

The sight of hundreds of young, enthusiastic researchers is awe-inspiring. But not all will contribute significantly to the battle, says Kevin Marsh, director of the Kenya Medical

Research Institute (KEMRI) in Nairobi. “If you look at the number who will be internationally competitive scientists, the figure remains low,” he says, pointing to the dearth of well-equipped, African-run laboratories as the main cause.

The lack of infrastructure and career ladders in Africa also means that the best students often emigrate, complains Nwaka. Mbacham was only able to stay in Cameroon thanks to funding from the Gates Malaria Partnership run by the London School of Hygiene and Tropical Medicine. Most who stay have to put up with very low salaries. “Paying highly qualified personnel around \$300 a month is an invitation to leave,” says Kwadwo Koram, a researcher at the Noguchi Memorial Institute for Medical Research in Accra, Ghana.

KEMRI, supported by UK medical charity the Wellcome Trust, is one of the few centres that can compete with international labs for good students, with state-of-the-art equipment and an army of top scientists.

Grass-roots initiative

Another jewel is the Malaria Research Training Centre in Bamako, Mali, supported by the US National Institutes of Health. “From three doctoral-level scientists at its creation in 1992, the centre has now over 30 MDs, PharmDs and PhDs,” boasts Abdoulaye Djimde, head of its epidemiology and immunology department. His team discovered the genetic markers for *Plasmodium falciparum* parasites resistant to the drug chloroquine (A. Djimde *et al.* *N. Engl. J. Med.* **344**, 257–263; 2001).

In Mozambique, the tin-roofed Manhica Health Research Centre shows what can be achieved in the face of adversity. Set up in 1996 when the country was still recovering from civil war, the centre is gathering data on 65,000 people, and runs a vaccine trial involving 2,000 children.

Despite such achievements, the research landscape is still largely bare in Africa and basic lab equipment is often non-existent. Nwaka spends much of his holidays teaching in his native Nigeria and he takes picnic cool-

ers loaded with reagents, hoping to restock local labs. But these often don’t have the money or a sufficiently reliable electricity supply to run a fridge.

The top centres that do exist all have strong ties to agencies in more developed countries, and this tends to take the power out of African hands, says Binka. “Show me one well-funded centre completely run by Africans,” he says. “Collaboration with the north is important, but we need Africans to start taking leadership.”

Getting connected

African scientists are increasingly organizing themselves into regional networks. These enable them to exchange ideas and skills, and pool their resources, says Wen Kilama, who heads the African Malaria Network Trust, a body based in Dar es Salaam, Tanzania, that trains scientists and supports vaccine trials.

Getting money for such networks from international programmes is difficult if not impossible, says Binka. Most don’t provide explicit funds for research or infrastructure in Africa. But there could be a windfall coming for African scientists — the European and Developing Countries Clinical Trials Partnership launched in 2002. This African-led partnership, funded by the European Union (EU), aims to put Africans in control of developing treatments for HIV/AIDS, tuberculosis and malaria by funding infrastructure and training at five or more sites.

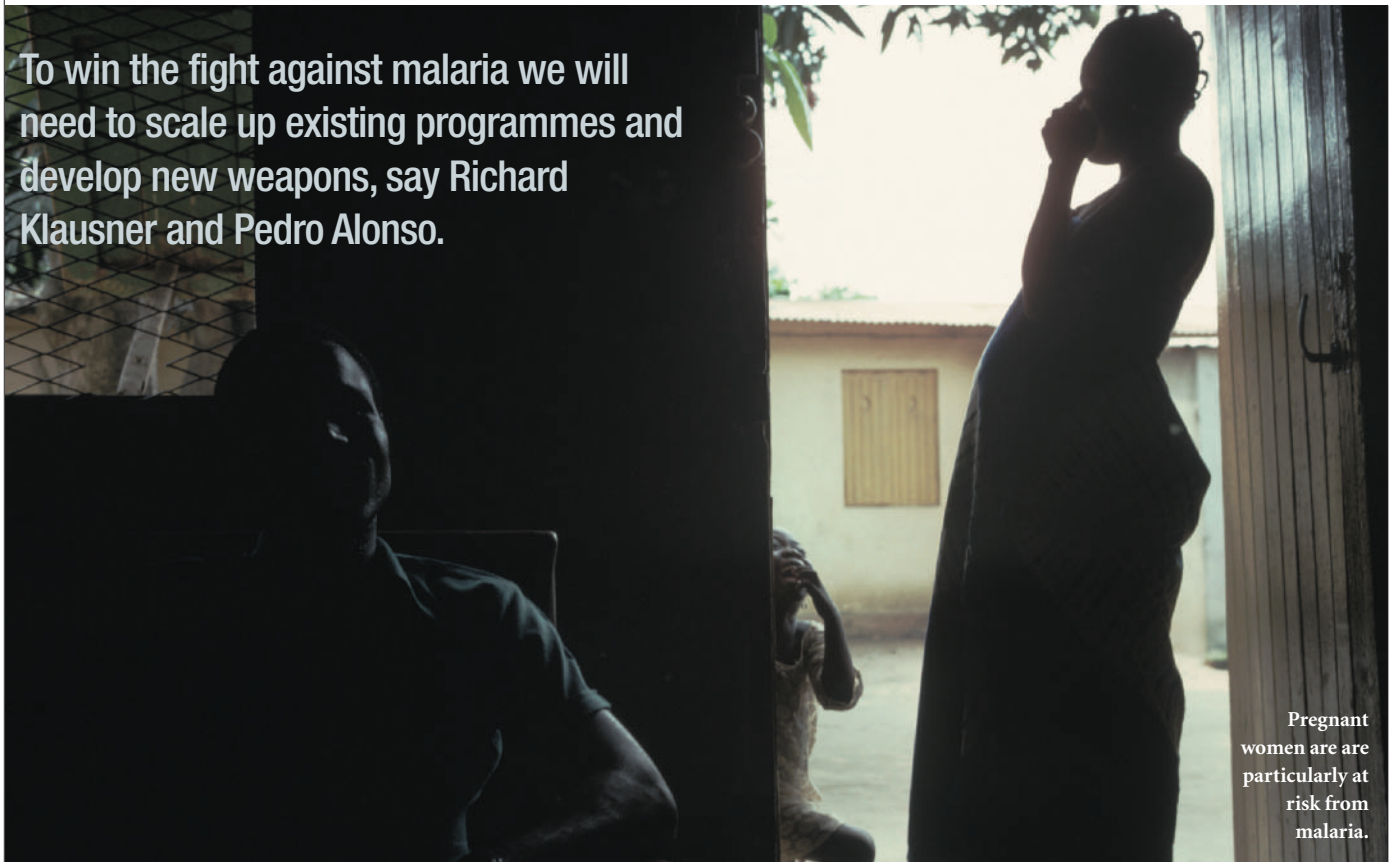
The EU has put €200 million (US\$245 million) into the scheme so far, and a further €200 million is expected from EU member states. Industry is expected to contribute a similar amount again. “It is certainly a major initiative, but it is too early to judge its success or failure,” says Bojang Kalifa, a clinician at the UK Medical Research Council’s laboratory in Banjul, Gambia.

If all that funding comes through, it will be a huge cash injection for African science. But much greater efforts will be needed before Nwaka can be confident of finding a working fridge — and before Malenga no longer needs to dread the storm clouds gathering over Blantyre’s peaks. ■

Declan Butler is *Nature’s* European correspondent.

An attack on all fronts

To win the fight against malaria we will need to scale up existing programmes and develop new weapons, say Richard Klausner and Pedro Alonso.



Pregnant women are particularly at risk from malaria.

The global malaria crisis is desperate and worsening, particularly in sub-Saharan Africa. Unless we find ways to control this devastating disease, efforts to reduce poverty, minimize childhood mortality, increase security and strengthen the most vulnerable societies will fail.

During the mid-twentieth century, the international community made decisive progress in its efforts to eradicate malaria from large swathes of the planet, but this momentum slowed during the 1960s with the realization that the available tools — indoor residual spraying with the insecticide DDT, followed by drug chemoprophylaxis — could not eradicate malaria in the regions with the highest rates of infection.

Only recently has there been a renewed political determination to tackle malaria, with initiatives and institutions being established to translate this commitment into action. These include the Roll Back Malaria initiative, the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the Malaria Vaccines Initiative. But these efforts are too fragmented and

short-term to succeed. The existing tools and strategies, although reasonably effective, can be greatly improved — if this doesn't happen, then they will get nowhere near eradicating the disease. The resources, policies and capacities to deliver them to those most in need also fall far short of what is required.

What is needed is determined action to deploy existing tools on a vastly greater scale, while investing in their improvement and the creation of new ones. Research must focus on tools that are suited to constraints in the field. Vaccines, for example, are well suited to resource-poor settings, and insecticide-treated bednets, particularly those with long-acting insecticides, are suitable for rural Africa. But the constraints are legion: lack of funds to purchase expensive drugs, the logistical complexity of sustaining labour-intensive indoor residual spraying of insecticides and, most importantly, the weakness of healthcare services in most endemic areas. Strive as we might to eliminate these constraints, the reality is that they are not going to disappear any time soon.

The renewed political commitment to tackle malaria is welcome, but the world now needs to go beyond 'calls to action' and declarations of spectacular targets for reducing the malaria burden — targets that will never be attained if we carry on with business as usual.

Three-pronged attack

Three major tools are currently used to combat malaria: controlling mosquitoes, reducing human–vector contact, and preventing and treating disease with drugs.

Vector control has saved millions of lives worldwide, through indoor residual spraying, environmental management to eliminate breeding sites, and use of mosquito larvicides. Spraying continues to play a major role in malaria control in much of Latin America and in parts of Asia. But its cost, logistical complexity and moderate efficacy make it poorly suited for controlling malaria in rural areas of sub-Saharan Africa.

Reduction of human–vector contact through insecticide-treated bednets is better

A. WEBB/MAGNUM PHOTOS

suitable for malaria control in Africa, enjoys greater community acceptance and is as efficacious as indoor residual spraying. Randomized trials of the nets in diverse settings have established their effectiveness at cutting malaria-related morbidity and mortality. For example, in parts of Africa where malaria is endemic, bednets reduce mortality from all causes among children under five by one-fifth for sustained periods¹. But although they are inexpensive and effective, fewer than 2% of Africans sleep under them². Massive campaigns to increase their use are required as a matter of urgency.

The cornerstone of malaria control worldwide remains effective and inexpensive drugs (see 'Between hope and a hard place', page 926). But resistance of the *Plasmodium* parasite to the most popular drug in Africa, chloroquine, is now widespread and few alternatives have been licensed over the past 20 years, or are in development.

Fresh strategies

Artemisinin derivatives offer great hope for reducing malaria morbidity and transmission. They can be particularly effective in combination with other antimalarials (artemisinin-based combined therapies or ACTs), which should also delay the development of resistance³. But their use has been limited by poor supply of high-quality artemisinin derivatives and by their high costs. These issues have led to vigorous debate about whether ACTs should become standard treatment for malaria in endemic countries. Even if supplies were unlimited, most African countries do not have the financial resources to purchase ACTs^{4,5}.

Besides treatment, malaria drugs also have the potential to prevent the disease and control its spread. Prophylaxis is highly effective for travellers to Africa, but in its present form has limited use among the populations of sub-Saharan Africa. Instead, new strategies for using antimalarials are being tested. For example, intermittent preventive treatment in infants (IPTi) uses existing drugs to protect infants from the worst effects of the disease. Infants receive an antimalarial three times during the first year of life at the time of routine immunization, whether or not they have malaria.

Two studies in Tanzania have shown that IPTi reduces malaria and anaemia in the first year of life by up to 60% (ref. 6). IPTi has the potential to become a major tool for malaria



Preventative measures: spraying with residual insecticides (top) can help to cut the number of hospital admissions with malaria.

control in Africa because it can be delivered through the Expanded Programme on Immunization (EPI), one of the best-functioning systems of regular health contact with young children in Africa.

The IPTi Consortium is an alliance of the World Health Organization (WHO), the UN Children's Fund (UNICEF) and leading research centres in Africa, Europe and the United States, with some US\$28 million of funding provided by the Bill & Melinda Gates Foundation. It has developed a research and implementation agenda that will rapidly resolve outstanding scientific questions about IPTi and move the intervention into policy and practice.

Long-term programme

Operating in parallel with this strategy is intermittent preventive treatment for pregnant women (IPTp). This approach has shown promise when the drugs are administered during antenatal clinic visits⁷, and although efficacy data are limited, it is now a recommended core intervention for improving malaria control during pregnancy. Additional research is needed to assess the safety and efficacy of different

malaria drugs during pregnancy and to evaluate the impact of combined interventions, such as insecticide-treated bednets and IPTp.

Even with plans to improve the use of existing tools, new drugs are urgently needed (see 'Winning the drugs war', page 942), as are financing and distribution mechanisms for their rapid introduction. As with other infectious diseases, prevention must be the mainstay of long-term efforts to control malaria, including the development and introduction of novel forms of environmental vector control (see 'Taking aim at mosquitoes', page 936) and an effective preventive vaccine (see 'Save the children', page 940).

Since the 1930s, it has been understood that "malaria control should not be a campaign, it should be a policy, a long-term programme"⁸. Yet the past four decades have seen few attempts to evaluate the impact of sustained malaria-control programmes that involve scaling up a combination of strategies over a large region, including appropriate treatment, prevention and vector control.

Now is the time to launch and evaluate complementary approaches to malaria control that cover a country or even an entire region of sub-Saharan Africa. Such projects could also generate the crucial evidence required to guide future campaigns. The Bill & Melinda Gates Foundation is currently in discussion with potential partners to assess the possibility of launching just such a research project.

Finally, the biggest challenge confronting the global community — with respect to both existing and future strategies — is to seize the enormous opportunities that already exist to reduce the malaria burden and improve global health. This will be a true test of global responsibility and solidarity. ■

Richard Klausner is executive director of the Global Health Program, Bill & Melinda Gates Foundation, Seattle, Washington. Pedro Alonso is at the Centro de Investigação em Saude da Manhica, Mozambique, and the Center for International Health, Hospital Clinic/IDIBAPS, Universidad de Barcelona, Spain.

1. Lengeler, C. *Cochrane Database Syst. Rev.* Issue 2, CD000363 (2004).
2. *Africa Malaria Report 2003* (WHO/UNICEF, Geneva, 2003); available at www.rbm.who.int/amd2003/amr2003/amr_toc.htm.
3. Adjui, M. *et al. Lancet* **363**, 9–17 (2004).
4. Attaran, A. *et al. Lancet* **363**, 237–240 (2004).
5. Butler, D. *Nature* **429**, 588 (2004).
6. Schellenberg, D. *et al. Lancet* **357**, 1471–1477 (2001).
7. Shulman, C. E. *et al. Lancet* **353**, 632–636 (1999).
8. Boyd, M. *Am. J. Trop. Med.* **19**, 1 (1939).

Where did it all go wrong?

International agencies have failed to meet their own malaria performance targets and should be held to account, says Amir Attaran.

In May 1998, the director-general of the World Health Organization (WHO) announced a United Nations-led campaign to Roll Back Malaria (RBM), pledging to halve malaria deaths by 2010. Today, RBM is at the halfway point and the WHO's own statistics show deaths have actually increased (see graph, right). For the several hundred thousand children who died in the interim, RBM is not just a failure but a fatal betrayal by the United Nations.

How did this happen? The fundamental reason is political: the WHO launched RBM to create an advocacy splash, but has subsequently failed to ensure adequate funding.

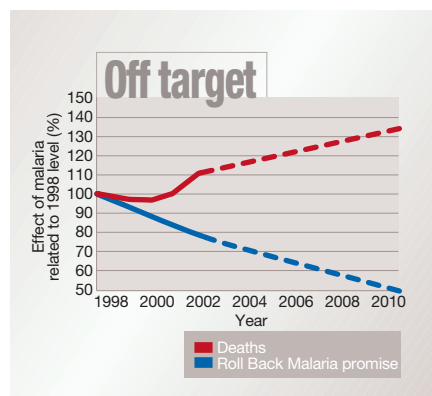
The world's foreign-aid donors — powerful developed countries and the World Bank — have behaved with their usual parsimony, but the WHO has not been vocal enough to hold them to account. Aid funding for malaria control has barely grown. No wonder the *British Medical Journal* recently called RBM “a failing global health campaign”¹. Without large increases in foreign aid, the least-developed countries that are most severely affected by malaria can do little. They have an average public-health expenditure of barely US\$6 per person annually².

This figure would be more had African governments honoured their promise at a conference in Abuja, Nigeria, in 2001 to spend 15% of their budgets on health. But even if they had, the vast majority of malaria control in the poorest countries would still need to be paid for out of foreign aid.

So how much foreign aid is available to control malaria? The WHO's most recent estimate was that “approximately US\$200



Roll Back Malaria's difficulties so far suggest that major aid donors such as the World Bank (pictured) are not rising to the challenge.



million was earmarked for malaria control worldwide” in 2002, including both domestic and foreign-aid budgets³. This is roughly consistent with an estimate for 2000 by myself and Vasant Narasimhan⁴ of \$100 million, which counted only the foreign-aid budget.

Most recently, the Global Fund to Fight AIDS, Tuberculosis and Malaria gave cause for optimism by pledging \$895 million over two years, rising perhaps to \$1.8 billion over five years⁵. At an average rate of \$360 million annually, assuming full disbursement, this will be the largest increase in malaria-control funding for decades⁶.

But making significant progress will require billions of dollars annually, not millions. There is a precedent: malaria-

control funding in the 1960s did reach such levels⁷. RBM has said that Africa needs \$1 billion annually³. A more careful peer-reviewed estimate is that at least \$1.6 billion is needed just for malaria medicines in Africa alone⁸.

Add in the cost of disease prevention — such as bednets and insecticides — training and staff costs, and the fact that many of the world's malaria fatalities occur outside Africa⁹, and a fair guess for the worldwide price tag must be at least \$5 billion annually.

Thus, even with the advent of the Global Fund, there is far too little foreign aid to combat malaria. This suggests that a more deliberate, more aggressive strategy to drive up donor aid is needed. To start with, it would be desirable to audit routinely how much each donor spends on malaria control, so that progress — or lack of it — can be tracked.

Tied up in red tape

At the moment, attempts to obtain foreign-aid figures are hampered by bureaucracy and poor availability of accurate, up-to-date data. Of 23 developed countries I surveyed in 2002, 13 seemed unable or unwilling to disclose their malaria-control funding, even after nine months and several reminders⁴.

For this article, I surveyed three major aid donors: the World Bank, the US Agency for International Development (USAID) and



The World Health Organization has pledged to halve malaria deaths by 2010 — but it has yet to make much progress.

the United Nations Children's Fund (UNICEF). The World Bank is the world's top donor of foreign aid, USAID is the foreign-aid agency of the world's only superpower, and UNICEF is the world's top child-health agency. How much does each spend on malaria?

A World Bank announcement on Africa Malaria Day in April 2000 pledged between \$300 million and \$500 million towards the eradication of malaria in Africa. In an earlier study, I estimated that the bank earmarked \$44 million of loans for malaria control aid in 2002 (ref. 4). For this article, the bank declined to supply me directly with data, but in a statement to *Nature* it claimed to have earmarked loans totalling between \$100 million and \$150 million for malaria control since 2000. Beyond that, it is difficult to know how much else the bank is spending on malaria. Its statement points to "non-earmarked monies" that may indirectly affect malaria (for example debt cancellation), but which it says are "difficult to quantify" because the bank does not track disease-specific details.

Whereas the World Bank supports a country's overall malaria-control budget, UNICEF and USAID tend to support discrete projects on the ground. Although RBM acknowledges that prompt and effective treatment are essential to reduce malaria deaths, neither UNICEF nor USAID buys more than

a tiny quantity of malaria medicines.

In personal correspondence, USAID said that it spends about 34% of its \$65.6-million annual budget for malaria on treatment. But it adds that it "typically does not purchase ... [malaria] medicines other than in exceptional or emergency circumstances", and that the quantities are "not large". It states that, instead, its strategy is "building the systems to procure, manage, and use the drugs".

Obsolete medicines

UNICEF does routinely buy and supply antimalarials, but the amounts are trivial: just \$3.7-million worth in 2003. Worse, most of that was spent on obsolete drugs that, because of resistance, usually do not work. Examples are chloroquine in Kenya and sulphadoxine-pyrimethamine in Burundi, which are so ineffective neither government sanctions their use. UNICEF spent only \$1 million on more effective drugs, the artemisinin combination therapies, even though these are the RBM-endorsed standard.

A similar picture emerges for buying and supplying bednets and insecticides. UNICEF spent \$17.3 million on these in 2003. USAID could not provide breakdowns for spending on each, although it spent \$8.4 million in 2003 on a partnership with private contractor NetMark, which sells (not gives) bednets to impoverished

Africans, a practice that is controversial.

UNICEF says that it also invests "notable sums" in programmes that strengthen healthcare delivery in general and which may benefit malaria control in the long run, such as the training of healthcare workers and education campaigns. But it provided no figures for what proportion of this investment directly relates to malaria control.

This all raises hard questions about legitimacy. The World Bank has said malaria slows economic growth in African countries by 1.3% per year. Over 35 years, this equates to a 32% reduction in GDP, worth in the range of \$100 billion¹⁰. If the bank fails to fulfil its own pledge on malaria control in Africa, while observing that its economies are so severely affected by the disease, is it actually doing international development? And if UNICEF, USAID and the World Bank cannot provide detailed audits of their malaria spending, is that a transparent use of taxpayers' money?

One lesson is clear: congresses, national parliaments and ultimately the public need to 'take back malaria'. Imagine if the managers of a project in a private company consistently missed their performance targets by a large margin. The company would be in trouble, and it would sack the managers.

Exactly the same should be true of the public agencies and officials who have for six years missed their performance target to reduce and ultimately halve malaria deaths. I suggest that legislators hold hearings into why the agencies have failed to roll back malaria, and take the necessary action. If that sounds harsh, weigh it against the number of children whose lives were tragically lost in the interim to a preventable, curable disease. ■

Amir Attaran is associate professor of population health and law at the University of Ottawa, Canada, and associate fellow at the Royal Institute of International Affairs, UK.

1. Yamey, G. *Br. Med. J.* **328**, 1086–1087 (2004).
2. *Macroeconomics and Health: Investing in Health for Economic Development* (WHO, Geneva, 2001).
3. *Africa Malaria Report 2003* (WHO/UNICEF, Geneva, 2003); available at www.rbm.who.int/amd2003/amr2003/amr_toc.htm
4. Narasimhan, V. & Attaran, A. *Malaria J.* **2**, 8 (2003).
5. www.theglobalfund.org/search/default.aspx?lang=en&component=Malaria (accessed 15/06/04).
6. www.theglobalfund.org/en/files/factsheets/disbursement.pdf (accessed 15/06/04).
7. Spielman, A., Kitron, U. & Pollack, R. *J. Med. Entomol.* **30**, 6–19 (1993).
8. Snow, R. W., Eckert, E., Teklehaimanot, A. *Trends Parasitol.* **19**, 363–369 (2003).
9. Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M. & Snow, R. W. *Lancet Infect. Dis.* **4**, 327–336 (2004).
10. McCarthy, F. D., Wolf, H. & Wu, Y. *Malaria and Growth* World Bank working paper 2303 (World Bank, Washington, 2000).



Under pressure: sub-Saharan Africa bears the brunt of malaria's scourge.

The invisible victims



The exact number of malaria sufferers in Africa is hard to quantify.

We need to know how bad the malaria situation is before we can make it better, says Robert Snow.

Despite the fanfare surrounding pledges to halve the number of malaria deaths over the next decade, the truth is that scientists have little idea how many people get sick or die from malaria, or where, let alone whether control efforts are working. Researchers lack data to evaluate accurately the impact of interventions such as drugs or insecticide-impregnated bednets. The international community must address this deficiency if it is to improve control strategies and measure their progress.

In Africa, the number of children dying of malaria has increased dramatically over the past 20 years, whereas other causes of childhood mortality have steeply declined¹. Some

90% of the world's malaria deaths probably occur in Africa. But many of our assertions about baseline data for Africa and elsewhere are couched in terms of 'probably' and 'likely', because the simple truth is we do not know.

The Roll Back Malaria (RBM) initiative aspires to halve malaria deaths by 2010 (ref. 2), and the United Nations' Millennium Development Goal is to halt and then reverse the rising incidence of malaria by 2015 (ref. 3). These targets are certainly optimistic, but are they even measurable?

National data on malaria illness and death are full of gaps and inaccuracies due to under-reporting and misdiagnosis. At the Kenya Medical Research Institute in Nairobi, we are instead using satellite data and climate models linked to demographic data to provide new scientific tools for predicting on a continent-wide scale where people live in relation to mosquitoes and malaria parasites^{4,5}.

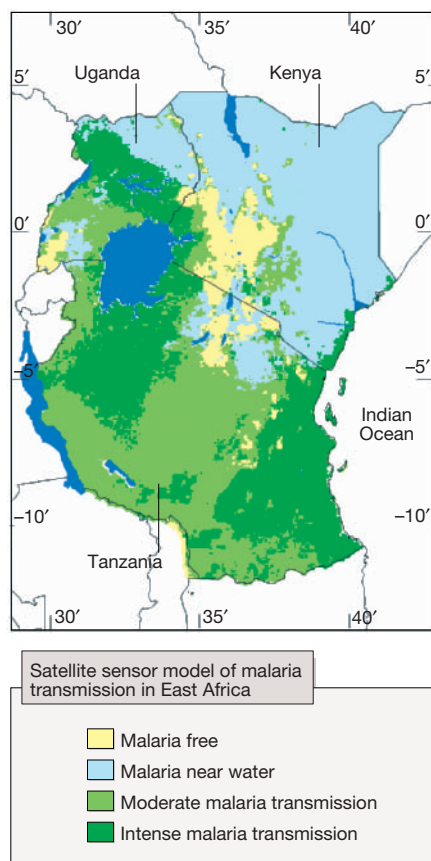
Combining data from epidemiological studies of malaria deaths and illness, this new science is helping RBM to overcome deficiencies

of national health information systems in Africa, providing more reliable estimates of the continent's *Plasmodium falciparum* burden⁶ (see map, opposite). But for other regions of the world, similar approaches have yet to be developed, or to be extended to other malaria parasites such as *P. vivax*.

RBM's decisions cannot be made effectively without accurate scientific data. What is needed is a properly funded, science-based strategy to evaluate the use of both existing and future control tools.

There are two relatively simple, effective and affordable interventions for malaria control: insecticide-treated bednets and artemisinin-based combination drugs.

RBM relies on national household sample surveys to monitor progress towards its goal of providing antimalarial drugs to 60% of febrile children by 2010. Current surveys report that 54% of febrile children in Africa receive an antimalarial drug⁷, but fail to capture information on the effectiveness, dose, timing or support of treatment. This is not surprising



given that the relevant questions are buried in a national-health survey questionnaire the size of the *Oxford English Dictionary*.

Similarly, RBM's target of providing bed-nets to 60% of children is not framed scientifically to ensure that these reach the most vulnerable. The distribution often correlates more closely with wealth than areas of high malaria risk⁷.

In 2001, the UK government granted one non-governmental organization US\$22 million to market nets in Kenya. But it focused on urban rather than rural areas. The outcome: a mere 1.5% increase in children protected—hardly a stunning success.

The best way to measure the impact of interventions would be to use rigorous, longitudinal demographic and epidemiological studies, such as those provided by the INDEPTH network of sentinel sites (www.INDEPTH-network.org). But such sites are few—there are only 11 in Africa. We urgently need a properly funded, science-based strategy to evaluate the use of both existing and future malaria control tools.

Clinical trials of new vaccines and drugs for malaria, such as those planned by the recently launched European and Developing Countries Clinical Trials Partnership, will eventually require sites where operational

effectiveness can be measured.

More money won't be enough. Donors, governments and international initiatives must take on board the concerns of scientists in the field, and back political goals with hard science. Without a greater appreciation of the importance of creditable baseline data, and the need for adequate funding and resources to gather such data, in a decade or so we will not be celebrating the rolling back of malaria—we will be scratching our heads wondering whether we made any difference at all.

Robert W. Snow is at the Kenya Medical Research Institute/Wellcome Trust Collaborative Programme in Nairobi, Kenya.

1. Snow, R. W., Trape, J. F. & Marsh, K. *Trends Parasitol.* **17**, 593–597 (2001).
2. *The Abuja Declaration on Roll Back Malaria in Africa* (2000); available at www.usaid.gov/our_work/global_health/id/malaria/publications/docs/abuja.pdf.
3. United Nations Development Programme *Human Development Report 2003. Millennium Development Goals: A Compact Among Nations to End Human Poverty* (Oxford Univ. Press, New York, 2003); available at hdr.undp.org/reports/global/2003.
4. Rogers, D. J., Randolph, S. E., Snow, R. W. & Hay, S. I. *Nature* **415**, 710–715 (2002).
5. Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M. & Snow, R. W. *Lancet Infect. Dis.* **4**, 327–336 (2004).
6. Snow, R. W., Craig, M. H., Newton, C. R. J. C. & Steketee, R. W. *The Public Health Burden of Plasmodium falciparum Malaria in Africa: Deriving the Numbers* DCCP Working Paper 11 (DCCP, Bethesda, Maryland, 2003); available at www.fic.nih.gov/dccp/wps/wp11.pdf.
7. Monasch, R. et al. *Am. J. Trop. Med. Hyg.* (in the press).

Struggling to make an impact

Hampered by bureaucracy, politics and ineffectual policies, critics claim that the international Roll Back Malaria (RBM) partnership is failing, and is a long way off its goal of halving malaria deaths by 2010. Few of the 44 African countries that signed up to its main goals, including providing immediate access to treatment for 60% of patients by 2005, have followed up with increased resources.

Allan Schapira, RBM coordinator at the World Health Organization (WHO), disputes that the programme is off-course, arguing that it will take years for the impact of current initiatives to become apparent. "There's no basis for the contention that deaths from malaria have gone up," he says, adding that tracking trends in malaria-specific deaths over short periods is difficult. What is clear, however, is that no data show substantial drops in deaths in sub-Saharan Africa, where most cases occur.

RBM was launched in 1998 by the WHO, the World Bank, the United Nations Development

Programme and UNICEF, and is now supported by 90 countries. It has raised the disease's international profile, says Roger Bate of the American Enterprise Institute for Public Policy Research, a Washington-based think-tank. "Unfortunately, I think that's kind of where the good news ends," he says. "If RBM were a government, it would be voted out of office."

In 2002, an external evaluation panel concluded that RBM was failing at regional and national levels, and was too isolated from local health policies (see *Nature* **419**, 422; 2002). It also found the World Bank's commitment to be ambivalent. "We are committed to doing a lot more, to doing much better," says Olusoji Adeyi, communicable-diseases coordinator at the bank.

RBM has since been reorganized to make it more accountable to malarial countries, and to have a tighter focus on fewer countries. Mark Young, UNICEF's senior health adviser for RBM, points to 'basket funding', where money is pooled nationally and health

ministries make spending decisions.

But critics also complain about the programme's inertia in replacing obsolete drugs with more effective, but costlier, artemisinin-based drugs (see *Nature* **429**, 588; 2004). In addition, they question its focus on mosquito nets soaked in insecticides to the exclusion of indoor spraying with DDT and other insecticides.

RBM claims that it promotes DDT wherever appropriate, but argues that nets are safer and more effective. Spraying is difficult to push through politically, says Schapira, with pressure from government and other donors. "We have had very, very strong lobbying over DDT," he says. "We have had to give up."

The biggest issue, says Schapira, is money. Most new funding for African countries is coming from the Global Fund to Fight AIDS, Tuberculosis and Malaria, although it is far short of the billions needed. "We have the tools and the strategies, but without more money it will be a disaster."

Apoorva Mandavilli is news editor of *Nature Medicine*.

Taking aim at mosquitoes

The malaria vector is back in scientists' sights, says Janet Hemingway, with insecticides and transgenic insects offering fresh hope.

No mosquitoes — no malaria. The insects are essential for transmitting the *Plasmodium* parasite, so for a century the mainstay of control was indoor residual spraying of insecticides. In the past decade, that has changed, and we have come to rely on antimalarial drugs as the main means of control. But with parasite resistance to cheap antimalarials growing, mosquito control is back on the agenda — and it will need new tools if it is to become a sustainable option.

Indoor spraying or bednets impregnated with pyrethroid insecticide can dramatically reduce death and illness, but a question mark hangs over their long-term effectiveness. Mosquito resistance to pyrethroids may hamper the use of impregnated bednets¹. Our understanding of the effect of control on the interaction between malaria transmission rates, acquired immunity and disease incidence is poor. Children exposed to infected mosquito bites can build up immunity and maintain it as adults if regularly exposed. But if transmission drops, they can lose immunity and risk developing severe malaria if it increases again.

So control methods that cut mortality in the short term could make populations more vulnerable in the long term. Encouragingly, a recent study in Kenya showed that the benefits of bednets remained six years after their introduction².

To make progress, existing methods will have to be deployed more effectively. In many countries malaria occurs mostly in the poorest, rural sectors of society, and even relatively simple control methods are rarely applied effectively. Fewer people competent in large-scale mosquito control are now working in Africa than 50 years ago, owing to the phasing out of national control programmes after malaria eradication campaigns failed in the 1950s and 1960s. To make



Making a splash: mosquito control plays an important part in reducing death and illness.

matters worse, insecticide resistance outpaces the appearance of new insecticides.

There are obvious ways to improve the way we tackle mosquitoes. Technologies such as remote sensing and climate modelling can map and monitor mosquito populations without the need for armies of insect collectors. Electronic data collection and transfer could be better used to feed information to control managers in real time. Such tools are starting to improve understanding of malaria transmission dynamics, but they need to be adapted to provide useful data at local scales and made available to affected communities.

Another exciting possibility is to exploit the mosquito and parasite genome sequences^{3,4} to develop insecticides that attack new targets within the vector or thwart resistance to existing pesticides. Such innovations could be in use within ten years.

Mosquito modification

Other genetic approaches include modifying mosquitoes to produce offspring that cannot transmit disease. Researchers have made great strides in this area. Introducing transgenics into the wild would not depend on health infrastructure. But where several species of vector are present, a separate transgenic must be created for each — a far from trivial undertaking. For example, it is still impossible to engineer *Anopheles funestus*, the major vector in southern Africa.

Alternatively we could use treatments based on RNAi (double-stranded RNA fragments that block the expression of specific genes) to 'silence' mosquito genes. Comparisons of the *Drosophila* and *Anopheles gambiae* genomes have identified several

mosquito gene families, such as C-type lectins, that have undergone recent expansion^{5,6} and are potential targets. Transcriptional analysis of *Plasmodium* and its vector during parasite development^{7,8} is pinpointing genes unique to mosquitoes, suggesting we could block mosquito–parasite interactions without affecting other insects. Progress is rapid, but ideal candidates have yet to be identified⁶.

Scaling up these technologies presents further challenges. Most at-risk populations have a poor understanding of malaria transmission, so acceptance of measures often depends more on how well biting insects are controlled than on potential reductions in malaria. Political and social acceptance of the release of parasite-resistant mosquitoes will likewise be needed.

Alternatively, there may be ways to exploit the finely tuned relationship that has evolved between parasite and vector. Simply altering the redox state of the insect's gut, through which the parasite must pass, may stop transmission⁹.

Although these openings give cause for optimism, they are only the first small steps along the difficult road to implementation. But such challenges should not diminish our determination to combat malaria.

Janet Hemingway is director of the Liverpool School of Tropical Medicine, Liverpool, UK.

1. Brooke, B. D. *et al.* *Bull. Entomol. Res.* **91**, 265–273 (2001).
2. Lindblade, K. A. *et al.* *J. Am. Med. Assoc.* **291**, 2571–2580 (2004).
3. Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
4. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
5. Ranson, H. *et al.* *Science* **298**, 179–181 (2002).
6. Osta, M. A., Christophides, G. K. & Kafatos, F. C. *Science* **303**, 2030–2032 (2004).
7. Abraham, E. G. *et al.* *J. Biol. Chem.* **279**, 5573–5580 (2004).
8. Srinivasan, P. *et al.* *J. Biol. Chem.* **279**, 5581–5587 (2004).
9. McCarroll, L. *et al.* *Nature* **407**, 961–962 (2000).



The long and winding road

Documentary makers can get as close to the war zones of disease as doctors and researchers — perhaps even closer. Julie Clayton and Declan Butler talk to Kevin Hull about his experiences.

Like many people on the receiving end of bad news, Kevin Hull rode waves of emotions as he discovered the malaria crisis. “When you start finding out about malaria you first get intrigued, then confused,” he says, “then you get outraged and depressed — and then angry.”

Hull is an award-winning film-maker with more than 20 documentaries and television series under his belt, including *The War Machine* — a TV series about the secretive weapons-development arm of Britain’s Ministry of Defence. In 2002, he turned his camera on the war against malaria.

Taking a break from final editing in the cutting room of the production company Films of Record in west London, Hull admits that he first became interested after being hoodwinked by headlines announcing that a malaria vaccine was now on the way. But he says that the film’s initial focus on the search for a vaccine soon shifted to address a larger question: why is malaria control failing?

The quest for answers took Hull to Gambia, Kenya, Tanzania, Florida, Britain, France and Switzerland. With the same insistent style as the maker of *Fahrenheit 9/11*, malaria’s Michael Moore grilled those fighting the disease, from remote villages in Africa, to top research labs and the corridors of power in health ministries, UN agencies and the World Health Organization.

The resulting film, *Fever Road*, pulls together three issues that are usually dealt with separately: the personal experience of



Life through a lens: Kevin Hull was shocked by the suffering he saw when he shot his malaria film.

African villagers; the scientists’ efforts to defeat the disease; and the political and economic issues shaping the international community’s response to it. With its powerful narrative, Hull hopes that his film will make decision-makers sit up and listen. “I don’t normally make films that are about increasing public awareness of a problem,” says Hull. “But sometimes you just feel very strongly about something.”

The 90-minute documentary is scheduled to hit screens later this autumn, starting with BBC Four television in Britain, and Public Broadcasting Service in the United States. Videos and DVDs will then be distributed as part of an international awareness campaign.

Vaccine upsets

Hull’s initial plan of tracking vaccine development led him to Oxford, where trials at the Centre for Clinical Vaccinology and Tropical Medicine were getting good results (S. J. McConkey *et al. Nature Med.* **9**, 729–735; 2003). The ‘prime-boost’ vaccine created by Adrian Hill and his team looked like a promising candidate. Classical vaccines that provoke antibody production cannot get at parasites hiding inside cells, but Hill’s vaccine is designed to trigger killer T immune cells that can destroy infected liver cells.

Hull’s crew met up with the researchers again during field trials of adults in Gambia at the end of the year. There they found that the vaccine gave little protection against

natural infection. “It’s not straightforward,” says Hull.

The long road towards a vaccine inspired Hull to use pictures of villagers walking along dusty roads as the central metaphor of his film. “Everyone is trudging up and down these roads for hours and hours trying to get to hospitals that don’t have any drugs,” says Hull. “Adrian in a way is on the same road, trying to get to a vaccine.”

A roller-coaster ride

The ups and downs of malaria-vaccine development are frankly portrayed in the film, and Hill applauds Hull’s efforts. “If he gets across to audiences that vaccine development is a long sequence of steps rather than a moment of brilliant discovery that cracks it, then that will have been useful,” Hill says.

Emerging from the labs, Hull tracked down the people at the sharp end of dealing with the disease. In Kenya, he secured behind-the-scenes access to researchers and officials struggling within an undeveloped health structure, and spent a month filming local doctors trying to quash a malaria outbreak in a remote part of the country.

What made Hull angry was watching doctors who had nothing to prescribe but drugs — often purchased by international agencies — that have been rendered useless by parasite resistance. “How can you justify giving drugs that don’t work to people who are dying? It’s really upsetting,” he says.

“I was outraged about the injustice of the situation and the complexity of the problem, in terms of linking the science, economics and politics,” says Hull. “Malaria is a disease that we can do things about, at every level of prevention and cure. But deaths have doubled over the past ten years, and projections suggest that the same will happen over the next 20, because of a massive level of neglect. This is a terrible problem that should be dealt with.” Hopefully, he adds, his film will take the first few steps towards doing just that. ■

Julie Clayton is a freelance writer based in Bristol, UK. Declan Butler is *Nature’s* European correspondent.

Strength in unity

The world must increase collaboration to meet the pressing need for a malaria vaccine, argue Carter Diggs, Sarah Ewart and Melinda Moree.

A wide range of public- and private-sector stakeholders currently drive the search for a malaria vaccine. This diversity is an asset because a range of imaginative approaches will be needed to produce a vaccine against a foe as formidable as the malaria parasite. But many of the institutions and individuals involved have different motivations. The prospect of a vaccine that could prevent the deaths of more than one million children every year should bring the community together around common goals.

In vaccine development, the private sector usually has a focusing effect, uniting the multiplicity of actors needed to develop, test, regulate, manufacture and introduce new vaccines. Unfortunately, this driving force has been absent in malaria vaccine development due to the perception that profits from an eventual product would be small. As a result, many of the efforts take place in government or academic laboratories, or in industry heavily subsidized by philanthropic organizations or governments.

Despite this, an increasing number of vaccine candidates are now moving into clinical trials and progressing through the product-development pipeline. A more coordinated effort to address the key scientific, technical and introduction issues is now needed to accelerate this progress and ensure that a vaccine, when available, will be taken up without delay by immunization programmes in disease-endemic countries.

The World Health Organization (WHO) lists some 75 malaria vaccines or vaccine concepts that are under development, most of them funded by a small number of agencies. The advantage of this loose structure is



Raising the game: technicians look after *Anopheles* mosquitoes bred for research.

that there is competition to pursue innovative and diverse approaches. The disadvantage is that there can be gaps in some parts of the research base and unnecessary repetition in others.

Although active networks exist within the community, there is no single driving force towards an effective, licensed vaccine, nor is there a global plan to maximize progress and minimize unnecessary duplication of work. To overcome this, the malaria vaccine community should strengthen its networks and increase collaboration to ensure that all of the diverse activities it is pursuing will lead to a successful malaria vaccine.

The community will face some difficult choices. The experience of vaccine research is that few promising candidates move through the pipeline to advanced development. With limited resources available for funding research, the community will have to decide which vaccine concepts to pursue. When making these decisions, the challenge will be to broaden institutional and individual interests to emphasize the global health consequences and make the most of available resources.

Cash shortfall

Funding has gradually increased in recent years. Total annual expenditures for malaria vaccine development worldwide rose from US\$42 million in 1999 to \$65 million in 2003. A single new donor, the Bill & Melinda Gates Foundation, can account for most of this new infusion of funding. At current levels, however, if a candidate in

phase II clinical trials demonstrated sufficient efficacy, there would be insufficient funding available to proceed to phase III trials.

In 2001, the Commission on Macroeconomics and Health called for \$1.5 billion to go to targeted research and development for new drugs, vaccines, diagnostics and intervention strategies against HIV/AIDS, malaria, tuberculosis, reproductive health and other health conditions of the poor. Increased funding would considerably accelerate progress, especially by enabling parallel efforts that currently have to take place serially. But in the midst of the clamour of competing priorities, few international organizations or donors have responded to this call and new resources have been difficult to attain.

Once the efficacy of a vaccine has been determined, further increases in expenditure will be necessary to fund final process development, pivotal trials, installation of a manufacturing capability and introduction of the product.

The decision over whether to assume such an effort will be approached differently by the various groups involved. A vaccine manufacturer would necessarily emphasize business considerations, including projected profits or losses.

Aid organizations and private donors, on the other hand, will no doubt assess the cost-effectiveness of a malaria vaccine compared with other control measures. Ideally, vaccines will prove to be so effective that new funding will be provided to introduce immunization as an additional modality

"Ideally, vaccines will prove to be so effective that new funding will be provided to introduce immunization in addition to pre-existing interventions."

along with pre-existing interventions. Otherwise, difficult decisions on which of the several alternatives offers the most promise will be necessary. Aid organizations and donors would also consider long-term sustainability issues, for example the likelihood that funds would be available for vaccine purchase and distribution. The development of new tools that offer promise for success in fighting malaria, however, should generate new funding.

Ministers of finance and health in malaria-endemic countries would have to make their decision on whether to fund the distribution and delivery of a new vaccine in the context of the myriad other health interventions competing for the same limited pot of funding.

A coordinated and coherent message from the malaria community as to the role and value of interventions against malaria, including a vaccine when available, will be



Mosquitoes feed on blood infected with *Plasmodium* as part of vaccine research.

critical to international and national-level policy-makers. An existing umbrella organization, the Roll Back Malaria Partnership, could strengthen the ties between the different players in the malaria community, present a coherent message and aid the deci-

sion-making process for policies on malaria control as new tools become available.

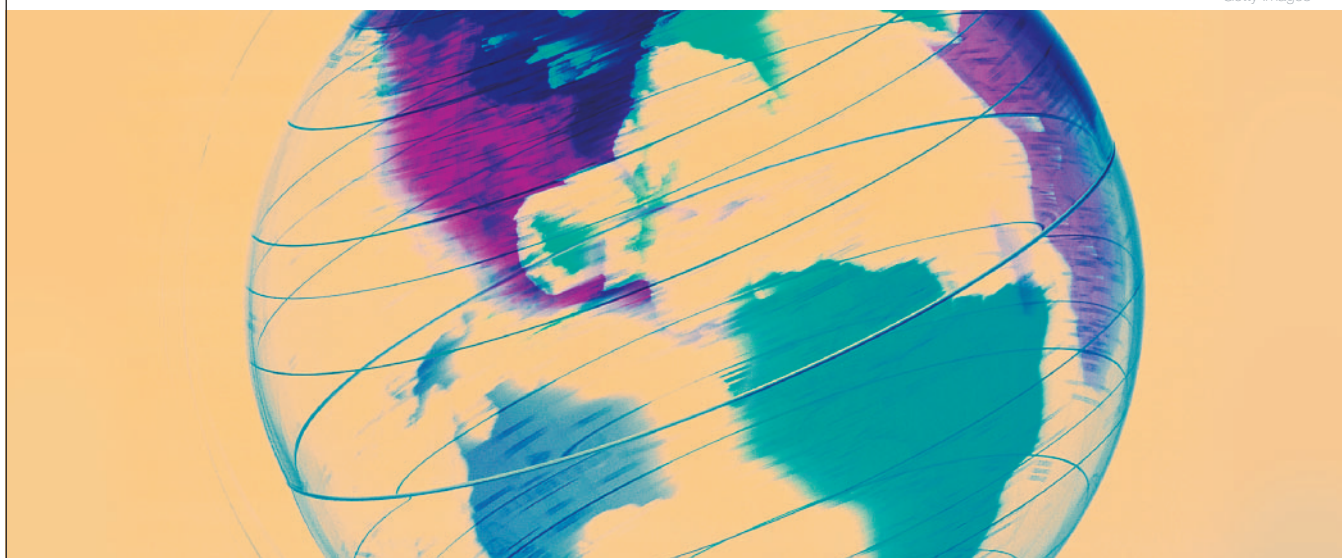
United against a common foe, those involved in research and development and those working to control the disease in the field could come together to tell the compelling story of the impact of malaria on the lives of people in developing countries and so motivate funders to do more to fight malaria. There can be little doubt that this coordinated approach towards major policy issues will accelerate the introduction of a malaria vaccine. Those responsible must vigorously take up the challenge. ■

Melinda Moree is director of, and Sarah Ewart a policy analyst at, the Malaria Vaccine Initiative. Carter Diggs is senior technical advisor at the USAID Malaria Vaccine Development Program.

The views expressed by the authors do not necessarily reflect the views of the US Agency for International Development and the US government generally.

Nature Publishing Group supports SCIDEV.NET, HINARI and AGORA to provide access for both readers and authors in the developing world

Getty Images



Find out more about Nature Publishing Group at www.nature.com

nature.com
the world's best science on your desktop

nature publishing group 

Save the children

Creating a malaria vaccine will be tough. But Africa needs one now more than ever, says Stephen Hoffman.

The malaria parasite *Plasmodium falciparum* kills more children under the age of five than any other infectious agent. The need for an effective vaccine is more urgent than ever: *Plasmodium* parasites are increasingly resistant to drugs, as are the *Anopheles* mosquito vectors to insecticides.

There are many successful vaccines against viruses, such as smallpox and polio, and bacteria, such as *Haemophilus influenzae*, but there are no commercially available vaccines for human parasites. These present a far greater challenge because they are more complex.

First, they have much larger genomes coding for more proteins. Second, they have multi-stage life cycles in which they express many different proteins at different times (see Figure, below). As a result, protective immune responses against the extracellular sporozoites that enter with the bite of a mosquito, for example, may have no effect on the asexual 'erythrocytic stage' merozoites that later emerge from the liver and infect red blood cells. Third, *P. falciparum* in particular has enormous variability in its proteins. This is critical to the parasite's survival, enabling it to evade host immune defences. It also means that a vaccine containing a single sequence of a single protein, or just a few, may fail to have a large, sustainable impact.

There is little precedent for a successful vaccine being created in any way other than

through the production of whole organisms, generally in laboratory culture, which for malaria has so far proved impossible. All commercially available vaccines but one consist of material from whole viruses or bacteria, or purified components. The only successful recombinant-protein vaccine is the hepatitis B surface-antigen vaccine. Almost all malaria vaccine candidates are based on individual components (generally proteins or parts of proteins) that have been created in the lab using recombinant proteins, synthetic peptides, recombinant viruses and bacteria, or DNA or RNA plasmids.

Finally, the most effective 'subunit' vaccine may need to induce both antibody and T-cell responses. Antibodies could block sporozoites as they enter the body, but have to act within minutes to block entry into the liver. They can also prevent infection of red blood cells, help destroy those already infected and prevent infection of mosquitoes. T cells have the potential to kill infected liver cells, thereby controlling and even eliminating infection. Both types of response may have to be directed against multiple proteins, at different stages of the life cycle, and at the same time. If so, vaccine developers face a technical problem that has never been solved.

Scientists who are attempting to develop an effective vaccine against malaria keep two observations in mind. The first is that most malaria deaths and severe disease in sub-

Saharan Africa occur in infants, young children and pregnant women. Adolescents and adults rarely develop severe disease or die after repeated infection with *P. falciparum*. They have presumably developed natural immunity that limits parasite replication and severe forms of malaria, but does not prevent infection resulting in milder symptoms. Pregnant women, especially with their first child, seem to lose this immunity. This observation has led to the idea that a vaccine would be worthwhile even if it only limited the severity of disease for those most at risk, without preventing infection or moderate disease. Such a vaccine would probably not be very useful for tourists, but would be beneficial in most parts of sub-Saharan Africa.

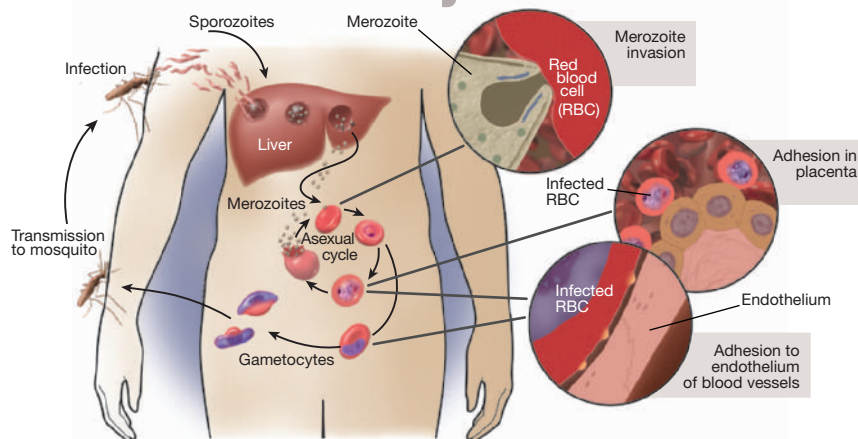
A thousand bites

The second observation is that when volunteers are exposed to more than a thousand bites from *P. falciparum*-infected *Anopheles* mosquitoes that have been irradiated to weaken the sporozoites they carry, they develop protective immunity against multiple strains of *P. falciparum*. If these volunteers are exposed to normal sporozoites, more than 93% are completely protected against developing erythrocytic stage infection¹. This is the strongest evidence that a highly effective vaccine is possible. So some vaccinologists, including this author, are focusing on achieving a more powerful vaccine that prevents all infections with *P. falciparum* in more than 85% of recipients².

Vaccinologists have adopted three main strategies. The first is to create vaccines that counter sporozoites as they enter the body and invade and reproduce in the liver (pre-erythrocytic stage vaccines). These have the potential to prevent infection altogether. The second is to limit invasion of erythrocytes and subsequent multiplication and pathological effects (asexual erythrocytic stage vaccines). Such vaccines would only limit severe disease — they would not prevent infection or mild disease. The third strategy is to prevent the spread of viable parasites to other people with 'transmission-blocking vaccines'. These stimulate the production of antibodies that are ingested when the parasite is sucked up by a mosquito. The antibodies destroy the parasite within the vector's gut.

Vaccinologists may need to combine all three strategies to have the best chance of success. Current vaccine candidates in clini-

Malaria life cycle





Children may benefit most from a vaccine — even if it only limited the severity of the disease.

cal trials, however, contain just one or a few proteins. In contrast, the protective immune responses elicited by natural exposure to malaria or by immunization with radiation-attenuated sporozoites could be directed at many, perhaps hundreds or even thousands, of the proteins encoded by the 5,300 genes in the *P. falciparum* genome.

On test

All three strategies are now being pursued, thanks to recent increases in funding. About US\$65 million was invested in malaria vaccine research in 2003, and some \$85 million is being invested in 2004. The US National Institutes of Health has provided the largest amount (\$33 million in 2003), with the Malaria Vaccine Initiative in second place, giving \$14 million in 2003. According to the World Health Organization, there are 25 candidates in phase Ia testing (safety and immunogenicity), six in phase IIa testing (efficacy against experimental challenge), eight in phase Ib testing (safety and immunogenicity in a disease-endemic country) and two in phase IIb testing (efficacy in a disease-endemic country).

Despite all of these efforts, only one *P. falciparum* protein, the circumsporozoite protein (PfCSP), has been repeatedly evaluated in clinical trials and shown to provide complete protection in a portion of volunteers. The lead candidate based on this protein is called RTS,S/AS02A, which GlaxoSmithKline Bio-

logicals and the Walter Reed Army Institute of Research in the United States initially developed, and which the Malaria Vaccine Initiative is now supporting in large-scale trials. In its first trial, the vaccine protected six out of seven volunteers against *P. falciparum* challenge three weeks after the last immunization³, but subsequent tests found that it protected only 40–50% of volunteers within 2–3 weeks⁴. It protected 70% of semi-immune Gambian adults for two months, but no longer⁵. RTS,S/AS02A is now being tested in a phase IIb study in 2,000 children in Mozambique, with results expected this autumn.

Delayed onset

Just one other vaccine candidate, developed at the University of Oxford, has reached phase IIb trials. It includes sequences from several pre-erythrocytic stage proteins, and the entire coding sequence of PfTRAP/PfSSP2. It is delivered using a 'heterologous prime boost' strategy that involves initial doses containing a DNA plasmid or recombinant pox virus expressing *P. falciparum* epitopes and whole proteins, followed by later injections of a recombinant modified vaccinia virus encoding the same proteins. This approach reproducibly delays the onset of parasitaemia in volunteers, but has prevented infection entirely in only a few recipients⁶.

With large numbers of candidates in preclinical and clinical development, many more, especially asexual erythrocytic stage

vaccines, are likely to enter field trials in the next five years. Furthermore, emerging genomic and proteomic studies of *P. falciparum*⁷ will lead to the development of even more candidate vaccines.

Until recently, most groups working on pre-erythrocytic subunit vaccines aimed for complete protection. But results from clinical trials have indicated that the current vaccines are unlikely to accomplish this in the majority of recipients^{3–6}. Many investigators are nonetheless continuing in the hope of at least achieving a product that limits severe disease. Others, including this author, have since re-examined the model system on which the strategy was based: the immunization of volunteers with radiation-attenuated *P. falciparum* sporozoites, followed by parasite challenge through the bites of infected mosquitoes⁸. We have founded a company, Sanaria, to develop a radiation-attenuated *P. falciparum* sporozoite vaccine. The manufacturing process is currently being optimized, with clinical trials planned for within 18–24 months.

When can we expect a vaccine to be widely available to infants in Africa? GlaxoSmithKline estimates that RTS,S/AS02A could be licensed as soon as 2010 (ref. 8 and personal communication). But gearing up to produce enough vaccine for widespread use, and having it deployed as part of the Expanded Programme for Immunization, will take longer. The time taken since the first PfCSP-based vaccine entered clinical trials in 1986 demonstrates how long and arduous the development process is.

There can be no doubt that an effective vaccine would have an enormous impact on the appalling toll of malaria. The dedication of many scientists and increased investment by government and non-profit institutions have led to the recent explosion of new candidates. We must work hard to sustain and increase both the enthusiasm and investment if we are to realize the dream of a vaccine for those whose lives are devastated by malaria. ■

Stephen Hoffman is chief executive and scientific officer of Sanaria in Rockville, Maryland.

- Hoffman, S. L. *et al.* *J. Infect. Dis.* **185**, 1155–1164 (2002).
- Luke, T. C. & Hoffman, S. L. *J. Exp. Biol.* **206**, 3803–3808 (2003).
- Stoute, J. A. *et al.* *N. Engl. J. Med.* **336**, 86–91 (1997).
- Kester, K. E. *et al.* *J. Infect. Dis.* **183**, 640–647 (2001).
- Bojang, K. A. *et al.* *Lancet* **358**, 1927–1934 (2001).
- McConkey, S. J. *et al.* *Nature Med.* **9**, 729–735 (2003).
- Hoffman, S. L., Subramanian, G. M., Collins, F. H. & Venter, J. C. *Nature* **415**, 702–709 (2002).
- Long, C. A. & Hoffman, S. L. *Science* **297**, 345–347 (2002).

Winning the drugs war

We have the science to make new antimalarials, say Robert Ridley and Yeya Toure, but we need better mechanisms and resources to develop drugs and deliver them.

In the 1990s, prospects for antimalarial chemotherapy became increasingly bleak owing to growing parasite resistance to chloroquine and sulphadoxine-pyrimethamine. Artemisinin-based drugs showed great promise in southeast Asia, but were barely used in Africa because of their cost and a lack of clinical data. To make matters worse, the pipeline for new drugs was dry as nearly all pharmaceutical companies had pulled out of antimalarial research.

The outlook for new antimalarials is now better than it has been for decades, thanks to public-private partnerships and increased funds for countries to buy drugs. But to deliver on this promise we must integrate drug development into a broader research agenda and involve stakeholders in developing countries.

Whereas the science is the fundamental obstacle to developing malaria vaccines, the main bottlenecks in drug discovery are structural. Resources are lacking to translate basic science into drug leads. In addition, mechanisms and resources are needed for improved clinical evaluation of drugs.

The brighter outlook for new drugs is in large part due to the nonprofit organization Medicines for Malaria Venture (MMV), established in late 1999 to fund and manage the discovery, development and registration of antimalarials. With a US\$20-million annual budget funded mainly by the public



sector and philanthropic foundations, it now runs about 20 drug-discovery and development projects. It has also persuaded the pharmaceutical industry to re-engage in malaria drug development¹.

Despite renewed international commitment to research and control, the reality for many patients is much as it was in the late 1990s. Children are dying because they are unable to access treatment and often the drugs they are given no longer work². Evidence from coordinated multicentre trials is now supporting a widespread rollout of artemisinin-based combination therapies in Africa³, but converting this policy into practice and implementing it on a nationwide scale is not straightforward. The new treatment regimens are more complex and expensive. Appropriate systems and finances must be put in place to ensure that artemisinin combinations are widely available and appropriately used to ensure their effectiveness and to prevent parasite resistance.

Productive partnerships

The current wave of antimalarial drug development has come from public-private partnerships. For example, registration of chlorproguanil-dapsone for the treatment of uncomplicated malaria resulted from a collaboration between GlaxoSmithKline and the Special Programme for Research and Training in Tropical Diseases (TDR)^{4,5}. Several fixed-dose artemisinin combinations are in late-stage development. Particularly noteworthy is the partnership between MMV and the Indian manufacturer Ranbaxy to develop a new class of molecule: a fully synthetic peroxide (see page 900)⁶ that may overcome several limitations of the artemisinins, including their short

half-life, short shelf-life and cost.

This project is a marvellous example of how a public-private partnership can take an innovative approach to drug discovery. With initial TDR support, chemists at the University of Nebraska had generated several novel molecules of interest. But it was not until industry partnered the project and MMV provided substantial funding and support combined with a rigorous review process⁷ that it could deliver a new drug-development target.

There is still a long way to go. Even if successfully developed and registered over the next four years, further research will be needed to assess just how the drug should be used in the fight against malaria.

Global support for such public-private partnerships remains insufficient. This is despite an annual increase of about US\$20 million in resources from public-sector organizations and philanthropic foundations such as the Bill & Melinda Gates Foundation, plus 'in kind' industry contributions such as GlaxoSmithKline's new malaria and tuberculosis drug-discovery unit in Tres Cantos, Spain¹. We need a significant scale-up of activity if we are to keep delivering new drugs.

We also need to research carefully how best to use such drugs. Development of fixed-dose combinations should be strongly encouraged to improve compliance and prevent the development of resistance. But such formulations have risks as well as benefits that need to be reviewed on a case-by-case basis, particularly for new drugs. The wrong choice of partner may limit the value of a new drug if it leads to problems in formulation, stability, toxicity or cost, or if resistance develops rapidly to the partner drug. Limitations on possible partner combinations may also



A shop in Kilifi, Kenya, sells malaria drugs and provides advice, although some therapies are now losing their effectiveness. But the potent antimalarial artemisinin (below) from the plant *Artemisia annua* (bottom) could provide an alternative treatment.



be imposed by the need for drugs to be available for niche applications, for example to treat severe disease or malaria in pregnancy, or for prophylaxis, including intermittent presumptive treatment in pregnancy and in infants (see 'An attack on all fronts', page 930). Decisions on whether and how to combine drugs may often need to be left until after initial regulatory approval of a single agent, or at least until late in development.

Taking the lead

To boost discovery of new lead compounds and exploit the potential of genomics (see 'Know thine enemy', page 944) we need innovations in administrative, managerial and funding systems as much as we need technical innovation⁷⁻⁹. We need to improve academics' access to high-throughput screening facilities in conjunction with parasite testing, exploratory chemistry and pharmacokinetics. Such approaches are now within the reach of some academic laboratories through non-proprietary chemical libraries and robotic technologies.

For some drug targets in the parasite, such as the cysteine proteinases and farnesyltransferases, the process can be simplified by 'piggybacking' on relevant industry-based chemistry to identify leads. A judicious mix of public-sector investment and public-private partnership needs to be further extended into this area.

Obtaining regulatory approval for a drug is just the first step. The efficacy and safety data required for approval, although critical, do not necessarily predict effectiveness in a real-life situation. For example, poor patient compliance with treatment regimens may have an impact on effectiveness, especially in settings where professional dispensing and medical supervision are lacking and patients are too poor to pay for a full course of treatment.

The thousand or so patients in a phase III clinical trial are too few to demonstrate fully a drug's safety. Rare adverse events occurring in sub-groups of patients — for example pregnant women, HIV-infected individuals, or patients taking other drugs — may not be noticed. Patients with different genetic susceptibilities, such as glucose-6-phosphate dehydrogenase (G6PD) deficiency, may also require special consideration. We need further post-registration studies on safety, patient compliance and effectiveness. We should also explore alternative formulations and improved drug-delivery strategies.

If properly managed, such studies would reduce the time it takes to decide whether

and how a drug should be included within national policies. We simply do not have the luxury of waiting many years for these data to accumulate in an ad hoc way. A more coordinated, comparative assessment of new drug options needs to take place in local settings to ensure best practice for rapid, timely and cost-effective treatment¹⁰.

Examples of such studies partnered by the TDR include assessment of rectal artesunate prior to hospital referral for patients unable to take oral medication, evaluation of chlorproguanil-dapsone in G6PD-deficient patients, and an assessment of Coartem in very young children. Attention is also turning towards the use of drugs during pregnancy.

We still fail to recognize the research and development role of scientists and organizations in developing countries. Agencies such as the US National Institutes of Health, the UK medical charity the Wellcome Trust and the new European and Developing Countries Clinical Trials Partnership are giving

researchers in developing countries a greater stake in clinical research, but there is still a long way to go. The wealth of opportunities provided by traditional medicines offers particular promise for developing capacity for drug research and development and we should build on

this through the Multilateral Initiative on Malaria and other organizations.

Although we now have considerably greater hope of developing a strong armamentarium of antimalarials, we still need to ensure that new drugs are optimally used and made accessible to patients. This will involve interactions across multiple disciplines, institutions and organizations. In particular, it demands coordinated action and dialogue between research and disease-control communities, both within countries and internationally.

Robert Ridley is director of the UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR). Yeya Toure is Malaria Research Coordinator at the TDR.

1. *Medicines for Malaria Venture Annual Report 2003* (MMV, Geneva, 2004); available at www.mmv.org/filesupld/184.pdf.
2. *Africa Malaria Report 2003* (WHO/UNICEF, Geneva, 2003); available at www.rbm.who.int/amd2003/amr2003/amr_toc.htm
3. Adjuik, M. *et al. Lancet* **363**, 9–17 (2004).
4. Lang, T. & Greenwood, B. *Lancet Infect. Dis.* **3**, 162–168 (2003).
5. Alleouche, A. *et al. Lancet* **363**, 1843–1848 (2004).
6. Vennerstrom, J. L. *et al. Nature* **430**, 900–904 (2004).
7. Nwaka, S. & Ridley, R. G. *Nature Rev. Drug Discov.* **2**, 919–928 (2003).
8. Ridley, R. G. *Nature* **415**, 686–693 (2002).
9. Fidock, D. A., Rosenthal, P. J., Croft, S. L., Brun, R. & Nwaka, S. *Nature Rev. Drug Discov.* **3**, 509–520 (2004).
10. *Scaling up Home-Based Management of Malaria: From Research to Implementation* (WHO, Geneva, 2004); available at www.who.int/tdr/publications/publications/pdf/home_2004.pdf.



Know thine enemy

The malaria and mosquito genomes will allow us to find new drug and vaccine targets, says Daniel Carucci.

The deciphering of the complete genome of the malaria parasite ushers in entirely new approaches to developing vaccines and drugs against this ancient pathogen. To conquer, or even control, malaria we need to fully understand the parasite's biology. Co-evolving with us over millennia, it has developed sophisticated tactics to survive our immune defences, and more than any other disease has spurred changes to the human genome itself.

The deadliest of the malaria parasites, *Plasmodium falciparum*, is thought to have originated more than 100,000 years ago, coinciding with the geographical expansion of humans out of Africa. We have been battling it ever since.

The parasite infects red blood cells, feeds on haemoglobin and eventually destroys the cells, often resulting in severe anaemia. It also makes the cells' surfaces sticky so that they cling to the lining of blood vessels, clump together and clog narrow arteries, blocking blood to vital organs including the brain. In the latter, 'cerebral malaria' can lead to coma and death.

The threat posed by malaria has been so great that, at the risk of potentially fatal sickle-cell anaemia in homozygous individuals, humans have evolved a mutant form of the haemoglobin gene that produces an

abnormally folded protein the parasite cannot digest. The fact that this deleterious mutation persists in our genome is a sign of the enormous pressure that malaria exerts over our evolution.

The continued success of the parasite—it infects more people today than ever before—is due in part to an ability to alter its surface proteins to deceive the host immune system, and to suppress immune responses to its quiescent liver stage (see 'Save the children', page 940). To develop vaccines and drugs that exploit vulnerabilities in the parasite's biology necessitates a complete understanding of the parasite and its complex relationships with its human and mosquito hosts.

Cracking the code

The genetic code of *P. falciparum*, published in 2002 (ref. 1), has yet to spawn extraordinary breakthroughs, but it has energized the malaria community and has also attracted a much broader range of scientists to join the effort. They have brought technologies such as gene chips, proteomics and comparative genomics. We now have the foundations necessary to study global gene and protein expression in the parasite, identify genetic polymorphisms, uncover mechanisms of drug resistance and elucidate the causes of severe disease.

The best way to capitalize on the malaria genome will be through partnerships and consortia that bring together bioinformaticians, biologists, vaccine and drug developers, and physicians working in endemic areas. These could provide a set of 'credentials' for each malaria gene, indicating, for example, what controls its temporal expression, variation between related species in populations and role in the parasite life cycle. This would allow researchers to prioritize targeted approaches to drug and vaccine development.

Gene chips have already been used to determine how each of the more than 5,400 parasite genes is turned on and off in the red-blood-cell stage of the life cycle^{2,3}, and to assign functions to particular proteins². We must discover how these genes evolved to counter human defences, and in particular those genes that the parasite is forced to alter to avoid the immune system and which represent potential vaccine targets.

The same gene-chip technologies that have been used to study genetic diversity in model systems such as yeast are helping to identify mutations in malaria parasite strains that may be associated with parasite virulence or those under immune selection⁴. Gene chips also have the potential to reveal how the parasite responds to drugs and to

identify mutants that confer resistance. This could help elucidate mechanisms of drug resistance and, ultimately, ways to combat it. So far, with only chloroquine tested, few changes have been identified at the RNA transcript level in response to this one drug. But changes in gene expression are likely to be seen with other drugs.

Pattern analysis

Proteomics — the study of the entire range of proteins produced by genes — is an exciting new field, and researchers are applying technologies developed in model organisms such as yeast to each stage in the life cycle of *Plasmodium*^{5,6}. Using micro-liquid chromatography and highly sensitive mass spectrometry, for example, scientists can catalogue the identity and location of the thousands of proteins present at each stage, including proteins expressed on the surface of infected red blood cells.

Patterns of protein expression can identify antigens that would make suitable targets for vaccines. For example, proteomics has been used to identify vaccine candidates in the sporozoite stage, by pinpointing the antigens that generate cellular immune responses in volunteers protected against infection by an attenuated sporozoite vaccine^{7,8}.

Plasmodium is one of the 'apicomplexan' parasites, which possess a vestigial chloroplast. But unlike a related parasite in this

group, *Toxoplasma gondii*, which can infect all eukaryotic cell types, *Plasmodium* is remarkable in that its different families are specific to particular vertebrate hosts, such as humans, mice, birds or lizards.

Understanding this specificity is key. New intervention strategies might be developed based on the mechanisms underlying this preference for different species and host cells. Why are human red blood cells susceptible to *P. falciparum* but not to related rodent malaria species? How do parasites recognize host and target cells?

Proteins that are differentially expressed



Plasmodium falciparum (yellow) emerges from red bloods that it has infected (above), while a researcher in Mozambique conducts research into how to combat the parasite.



"Our malaria adversary has evolved an impressive array of defences. Genomics will be critical to breaching them."

on the surface of *P. falciparum* sporozoites and merozoites, for example, enable invasion of liver cells or red blood cells,

respectively. But despite much research in the pre-genomics era, only a handful of molecules potentially involved in invasion are known, including those in the reticulocyte-binding homologue family⁹ that allow merozoites spewing from the liver to recognize red blood cells.

Comparative genomics may shed light on parasite–host interactions¹⁰ and lead to entirely new ways to interrupt crucial steps in the parasite life-cycle, perhaps using small molecules. Combining comparative genomics with gene-expression studies may identify new targets for drugs lethal to the parasite but non-toxic to humans.

In addition, it may be possible to extend the useful life of drugs by pinpointing drug resistance mechanisms, making the high costs of drug development a better investment. Although a daunting challenge, it may also be possible to investigate at the genomic level, in combination with case-control field trials, the role of certain malaria parasites in causing severe anaemia and fatal neurological complications.

Within the next five years, the sequences of the remaining five species of parasite will be completed, as well as those of several more mosquito vectors. This should help to identify the key molecules making the parasite specific to its host, and possible Achilles' heels.

But doing molecular genetics in the parasite is challenging because of the difficulties in culturing its various stages, and large-scale gene knockout experiments are fraught with formidable technical challenges. Our malaria adversary has evolved an impressive array of defences. Genomics will be critical to breaching them, but it's going to be a long and hard campaign.

Daniel Carucci is director of Grand Challenges in Global Health at the Foundation for the National Institutes of Health, Bethesda, Maryland.

1. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
2. Le Roch, K. G. *et al.* *Science* **301**, 1503–1508 (2003).
3. Bozdech, Z. *et al.* *PLoS Biol.* **1**, E5 (2003).
4. Volkman, S. K. *et al.* *Science* **298**, 216–218 (2002).
5. Florens, L. *et al.* *Nature* **419**, 520–526 (2002).
6. Lasonder, E. *et al.* *Nature* **419**, 537–542 (2002).
7. Doolan, D. L. *et al.* *J. Exp. Biol.* **206**, 3789–3802 (2003).
8. Doolan, D. L. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 9952–9957 (2003).
9. Duraisingh, M. T. *et al.* *EMBO J.* **22**, 1047–1057 (2003).
10. Cowman, A. F. & Crabb, B. S. *Science* **298**, 126–128 (2002).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: r.asami@naturejp.com

naturejobs

Granting longevity

Aaron Marcus has reason to celebrate. Last month, the physician scientist got his grant renewed for another five years — on work he started in 1955. The renewal makes him one of the longest continuously funded recipients of a grant from the US National Institutes of Health.

Marcus received his initial funding at the age of 29, while he was at the Veterans Affairs New York Harbor Healthcare Hospital and the Weil Medical College of Cornell University in New York. The grant was for a fellowship that would help him to isolate and characterize the role played by platelet lipids in blood clotting. He admits that it is a little bit “old school” to stick to one research theme for nearly 50 years. But by doing so, he has teased apart many particles and processes involved in clotting. His aim has been to help reduce the number of fatalities caused by thrombosis, which, he says, leads to a death in the developed world every 30 seconds.

The key to getting and keeping his grants — and making progress in untangling the process of clotting — has to do with asking very direct questions of himself and of his scientific goals, according to Marcus. When he teaches grant-writing, he says, he asks his protégés: “What are you gonna do? How are you gonna do it? Is it new, interesting and important?” Born in Brooklyn, New York, he retains a trace of his accent. “If you can answer those three questions, with quality, you’ll get the grant.”

These rules haven’t changed since he wrote his first grant, Marcus says. What has changed is that you need to show, early on, how the basic science could translate into medicine. “When you write a grant now, you need to show that it’s relevant to public health,” he says. Although he can’t guarantee that doing so means you will be funded for 50 years, it may be good enough for the first five.

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Forging a link between
chemistry and biology p948

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Breathing *life* into chemistry

Chemical biology, using chemical tools to solve biological problems, is awakening interest among students and creating a new breed of researcher, says Tim Chapman.

Theory and practice: taking notes from Harold Varmus (left); loading acrylamide gels (right).

CORNELL UNIV.



A. RAYNER

This summer, a small group of chemistry graduates at Cornell University went off to study biology at research laboratories in Manhattan. The students are the fourth group to enter an innovative multi-institution programme to produce scientists skilled in chemical biology, a cross-disciplinary field that promises to create a new breed of biomedical researcher.

The philosophy of the scheme, the Tri-Institutional Training Program in Chemical Biology, is to attract chemistry students into biomedical science. Although that was a popular move in the 1950s and 1960s, says programme director Timothy Ryan, it became less appealing as biochemistry morphed into molecular biology. In the post-genomic era, with researchers taking a quantitative approach to areas such as protein and cell biology, biologists are finding that they again need support from the physical sciences.

"Biomedical science really needs people with that training," Ryan notes. "Chemistry has some very nice niches and some very useful tools, which can be applied to some interesting areas of biology."

The challenge for training institutions is that few graduates with a physics and chemistry background want to commit themselves to a biology programme. They think they won't fit into the different culture, or would lose the benefits of their previous training. The idea behind the tri-institutional programme is to let them remain in a chemistry atmosphere while giving them as many biological problems to solve as possible.

The programme is a collaboration by Cornell's

chemistry department at Ithaca and three institutions based in New York city: Cornell's Weill Medical College, Rockefeller University and the Memorial Sloan-Kettering Cancer Center. Students alternate laboratory rotations and seminars in the city with advanced chemistry courses at Ithaca, plus an intensive course in biomedical science, before joining the laboratory of their chosen research adviser at any of the four centres. The rotation gives them a mix of basic chemistry, basic biomedical research and clinical research — all important components in the drug-discovery and drug-development pipeline.

TEN-YEAR VIEW

"We try to make new ties between the chemistry department at Ithaca and the biomedicine community down here," says Ryan, a biochemistry professor at Weill. "One of the connectors is the students, and we try to teach them as much biology as we can. Even if we don't capture them now, when they move on to their next position they'll be more attuned to working on biologically relevant problems. The success will be measured in maybe ten years' time when we see if we've placed into the research community a new breed of people with this hybrid training."

Harold Varmus, president of the Memorial Sloan-Kettering Cancer Center, says that there is a definite trend towards such cross-disciplinary projects. "All science is becoming much more interactive, especially at the intersection of chemistry and biology. We're

focusing on that when building graduate programmes, so that people actually get trained in both disciplines, as opposed to learning one as a student and trying to obtain the other later on," he says.

The Cornell programme is one of a new generation of interdisciplinary graduate courses in chemical biology — the use of chemical tools and principles to solve problems in biology. Demand is driven by the need to make the most of information coming from genomic and proteomic research. The courses aim to deepen understanding of protein structure and chemistry and of the role such structures play in cell functions.

In the United States, the National Institutes of Health has made chemical biology central to its 'road-map' for medical research. One of the document's main themes is that progress in medicine will require quantitative knowledge about the interconnected networks of molecules that comprise cells and tissues, along with improved insights into how these networks are regulated and how they interact.

Chemical skills and approaches will be invaluable in identifying and synthesizing small molecules to use as probes in studying the properties and interactions of proteins and ligands, as well as in designing and synthesizing new therapeutic compounds.

CHEMICAL TOOLS FOR BIOLOGICAL PROBLEMS

"The principal tools are those of synthesis, analysis, structure–function determination and mechanistic investigation," says Lawrence Marnett, director of the Vanderbilt Institute of Chemical Biology (VICB) in Nashville, Tennessee. Learning to use these chemical tools on biological problems will offer invaluable perspectives on a range of areas including drug discovery, structural biology and signal transduction.

The VICB, established in 2002 by Vanderbilt University's College of Arts and Science and School of Medicine, supports students from the different departments through three separate graduate programmes. "These provide students with an exceptional diversity of opportunities for pedagogical and laboratory training," notes Marnett. "An increasing number of students come from strong basic-science undergraduate programmes and want to apply their skills to important biomedical research problems."

Many universities now offer graduate programmes in chemical biology (see Web links). Although most take only chemical graduates — believing it is too hard for a biologist to make the leap to chemistry — others will take talented students from the life sciences.

The emergence of chemical biology is also reinvigorating chemistry. This is important in places such as the United Kingdom, where university courses in such traditional fields are often undersubscribed.

"We need to make these disciplines more attractive by going into areas such as chemical biology and nanotechnology," says Hagan Bayley, who last year became the first professor of chemical biology at the University of Oxford. "Many of these things exist already, but often you're just putting new labels on them so that people can recognize them, and to make these areas more attractive to students and to funding."

Bayley heads what he calls a "sub-department" within the chemistry department. Chemical-biology teams occupy a whole floor of the university's new £60-million (US\$110-million) Chemistry Research



Lawrence Marnett sees an exceptional diversity of opportunities.

Laboratory, and he is encouraging interactions across projects and disciplines.

"People are tending to do research more at the interface of different traditional disciplines, and that has created a need for undergraduate and graduate programmes in those areas," says Bayley. Such programmes will be developed over the next few years.

In Britain, chemical biology has been identified as a priority area by the Biotechnology and Biological Sciences Research Council, with the emphasis on protein chemistry and exploiting genomic data. The Engineering and Physical Sciences Research Council and the Medical Research Council are jointly soliciting research proposals for new applications of chemical tools to biomedical research. Several universities, including Newcastle upon Tyne and Warwick, already offer undergraduate degrees in chemical biology.

Ireland boasts the Centre for Synthesis and Chemical Biology (CSCB), a government-backed collaboration between University College Dublin, Trinity College Dublin and the Royal College of Surgeons in Ireland. New laboratories are being built for the centre, attached to University College's chemistry department. The CSCB has a student-exchange programme with the Centre for Medicinal Chemistry at the University of Regensburg, Germany, and is establishing links in Asia.

"We take the big view of 'molecules to medicine'," says director Pat Guiry. "It's not just synthesis for the sake of synthesis — there's a lot of collaboration with biologists where chemists can make suggestions about how to change the molecule to change its effect *in vivo*."

With 40 principal investigators and about 55 postdocs, the centre has some 120 PhD students, most with a chemistry background. "In some cases people who've done a chemistry degree are moving to biochemistry, but there's not as much movement as one would wish to see," says Guiry. "There's a culture change in how things are done, and people are starting to look outside the box. This is one of the centres helping to promote that sort of attitude."

The true test of this training will be when the current students emerge into the mainstream of research and apply their skills to some of biomedicine's most pressing questions.

Tim Chapman is a freelance writer based in Halifax, UK.

VANDERBILT UNIV.

Web links

Cornell Tri-Institutional Training Program in Chemical Biology

♦ www.med.cornell.edu/tpcb

Vanderbilt Institute of Chemical Biology

♦ www.vanderbilt.edu/vicb

Harvard Molecular, Cellular and Chemical Biology training programme

♦ mccb.harvard.edu

University of California, Berkeley, Chemical Biology graduate programme

♦ cbgp.cchem.berkeley.edu

University of California, San Francisco, Chemistry and Chemical Biology graduate programme

♦ www.ucsf.edu/ccb

University of Michigan Chemical Biology Interface training programme

♦ www.umich.edu/~chembio

Skaggs Institute for Chemical Biology

♦ www.scripps.edu/skaggs

University of Oxford Department of Chemistry

♦ www.chem.ox.ac.uk

University College Dublin Centre for Synthesis and Chemical Biology

♦ chemistry.ucd.ie/cscb